DOCUMENT RESUME

ED 044 322	24	SO 000 276

AUTHOR	Sykes, Gresham M.; And Others
TITLE	Law and Social Science Research, A Collection of Annotated Readings. Volume I.
INSTITUTION	Denver Univ., Colo. Coll. of Law.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C.
BUREAU NO	BR-6-1236
PUB DATE	69
GRANT	CEG-4-7-061236-0093
NOTE	432p.
EDRS PRICE	EDRS Price MF-$1.75 HC-$21.70
DESCRIPTORS	Curriculum Development, *Graduate Study, *Instructional Materials, Interdisciplinary Approach, *Law Instruction, Law Schools, Projects, Reading Materials, *Research Methodology, Social Sciences, *Sociology

ABSTRACT

This collection of readings, comments, and class notes has been prepared for the law student as an introduction to the research methods of sociology. The fundamental argument is, that in the growing use of social science by the law, the lawyer needs to know how the social sciences accumulate their evidence and build their theories. The materials in this collection are designed to cover two yearly quarters, with the first spent on the readings and exercises, and the second devoted to independent research projects by the law students. The sections on law and the social sciences, law and empirical inquiry, the design of research, and the collection of data include books, journal articles, and cases. The last four sections of this collection are in Volume II, SO 000 277. SO 000 275 describes the activities and objectives of the project. (SBE)

BR 6-1236
PA 24

U.S. DEPARTMENT OF HEALTH, EDUCATION
& WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ED0 44322

LAW AND SOCIAL SCIENCE RESEARCH
A Collection of Annotated Readings

By

Gresham M. Sykes
Robert H. Sulnick
Norman K. Linton

COLLEGE OF LAW

UNIVERSITY OF DENVER

(Prepared with the assistance
of a grant from
The U.S. Office of Education
#OEG-4-7-061236-0093)

SO 000 276

TABLE OF CONTENTS

INTRODUCTION

Editors' Comments

Readings:

Stuart Nagel, "Law and the Social Sciences: What Can Social Science Contribute?," *American Bar Association Journal*, 51 (1965), pp. 356-358.

Edmond Cahn, "Jurisprudence," *New York University Law Review*, 30 (1955), pp. 150-169.

Kenneth B. Clark, "The Desegregation Cases: Criticism and the Social Scientist's Role," *Villanova Law Review*, 5 (1959), pp. 224-235.

CHAPTER ONE: LAW AND EMPIRICAL INQUIRY

Editors' Comments

Readings:

Jacob Bronowski, *Science and Human Values* (Englewood Cliffs: Prentice-Hall, Inc., 1966), pp. 3-18.

Claire Selltiz, *et al.*, *Research Methods in Social Relations* (New York: Holt, Rinehart, and Winston, Inc., 1964), pp. 42-47.

Carl Hempel, *Philosophy of Natural Science* (Englewood Cliffs: Prentice-Hall, Inc., 1966), pp. 3-18.

Harry Kalven, Jr. and Hans Zeisel, *The American Jury* (Boston: Little, Brown and Company, 1966), pp. 3-11.

Jerome E. Carlin, *Lawyers' Ethics: A Survey of the New York Bar* (New York: Russell Sage Foundation, 1966), pp. 3-9.

William J. Goode and Paul K. Hatt, *Methods in Social Science Research* (New York: McGraw-Hill, 1952), pp. 68-73.

Class Notes

CHAPTER TWO: THE DESIGN OF RESEARCH

Editors' Comments

Readings:

Claire Selltiz, et al., Research Methods, pp. 8-23.

Samuel Stouffer, "Some Observations on Study Design," The American Journal of Sociology, 55 (January, 1950), pp. 356-359.

Jerome H. Skolnick, Justice Without Trial (New York: John Wiley and Sons, 1967), pp. 6; 23-37.

Roger B. Hunting and Gloria S. Neuwirth, Who Sues in New York City? (New York: Columbia University Press, 1962), pp. 1-7.

Marvin E. Wolfgang, deposition in case of Alabama v. Wheeler Billingsley, State of Alabama for Etowah County (Cases 1159, 1160, 1161).

Julius A. Roth, "Hired Hand Research," American Sociologist, (August, 1966), pp. 190-196.

Class Notes

CHAPTER THREE: THE COLLECTION OF DATA

Editors' Comments

Readings:

Claire Selltiz, et al., Research Methods, pp. 236-268.

William J. Goode and Paul K. Hatt, Methods in Social Research (New York: McGraw-Hill, 1952), pp. 170-183.

Howard S. Becker, "Problems of Inference and Proof in Participant Observation," American Sociological Review, 23 (December, 1958), pp. 652-660.

Charles T. McCormick, Evidence (New York: West Publishing Company, 1954), Excerpts from pp. 8-18.

"Cross-Examination of Police Officers," Trials, V, pp. 454-458.

Class Notes

INTRODUCTION

I

This collection of readings, comments, and class notes has been prepared for the law student as an introduction to the research methods of sociology.

Efforts to join law and the social sciences have been made in the past, of course, particularly in the areas of sociological jurisprudence and legal realism.[1] That body of work, however, appears to have been of greatest interest to legal philosophers and sociologists concerned with the general role of law in society. Only in recent decades has the attempt to build close bonds between the everyday work of the law and the social sciences begun to have a wide-ranging impact.

At the present time, law and the social sciences touch most closely at three main points. First, the practicing lawyer is confronted with evidence drawn from the social sciences in cases involving a change of venue, trademark infringement, discrimination, and so on. The scope of this material that is introduced as a part of litigation is as broad as the law itself and there are many indications that the findings of the social sciences will be even more important in future court cases and legal disputes than they are now. Second, the need for rational social planning and enlightened public policy has become ever more apparent in our modern, technological society. Legislation, as one part of that development, is more dependent than ever on scientific inquiry--including inquiry into social behavior. And the lawyer (who often plays a major role in the formation of public policy) is called on to assess the knowledge gleaned from many different disciplines, especially from the social sciences. Third, legal research is rapidly becoming something far different than a dredging up of cases, precedents, and statutory interpretations. Lawyers and legal scholars are devoting more and more of their attention to how the law works in reality. The study of law is no longer merely the study of what the legal rules should be or what the rules are--instead, there is a growing concern with the determinants of law and with what the legal rules accomplish in fact, and this is a matter that requires empirical investigation rather than a series of assumptions or reliance or dogma.

---

[1] See Bernard F. Cataldo, Introduction to Law and the Legal Process (N.Y.: John Wiley & Sons, Inc., 1965), Chapter I.

II

It is these three significant points of contact between law and the social sciences which are briefly reviewed by Professor Stuart Nagel in the first selection of readings. It is important to note, however, that

Professor Nagel's analysis of the contributions of the social sciences to the law is not exhaustive. There is at least one additional area in which the law is enhanced by the social sciences, namely, in the teaching of law.

All too often, legal rules, standards, and procedures are presented to the student in a vacuum. That is to say, the study of law is apt to be confined largely to discussions of appellate court decisions and various treatises with little attention paid to the social reality in which the law is embedded. But law is not a science unto itself--it is a body of rules, procedures, and remedies designed to cope with and regulate the actualities of human behavior. An understanding of the law, therefore, must be gained by examining the law in light of the existing state of the society and what we wish to achieve in the name of a just social order.

We do not, however, want to convey the idea that because law and the social sciences now show many signs of mutual interest and shared concerns, all is sweetness and light. There are, in fact, profound differences between the two fields which generate a good deal of antagonism and frequent misunderstanding. The law, it has been said, is organized for action; the social sciences, existing largely as an academic effort, are organized for accumulation of knowledge. The lawyer, then, or the law student or law professor, with an eye fixed on the practicing profession or the fine points of legal logic, is sometimes apt to view the findings of the social sciences as irrelevant or as a bothersome hindrance in the search for the applicable rule. The social scientist is, at times, no less suspicious, for a knowledge of the law and its operation is frequently missing from the intellectual baggage of even the well educated. As Max Radin has said:

> "An educated man feels that he must have in mind a modicum of information about history, science, economics, politics, art, and literature. He may make no pretensions--he generally makes no pretension--to be an expert in any of these fields, but he would be a little ashamed if he was crassly ignorant of them. But if it's a matter of law, the ordinary citizen is likely not only to be completely ignorant but to boast of his ignorance. He thinks of the law as something mysterious, irrational, and unnecessary. He has toward it a mixture of fear, resentment, and suspicion which he tries to cover by an attitude of almost constant derision."[2]

This mutual hostility is unfortunate, for although there are differences between the approach of the lawyer and the approach of the social scientist, there is also a shared interest in the realities of social life which man attempts to shape with social rules. A rather jealous concern for professional boundaries, a different degree of enthusiasm for specifics versus generalities, a varying taste for swift conclusions versus the endless questioning of science--all these may separate the lawyer and the social scientist; but their common concern with social norms, conformity, and rule-breaking behavior gives them a strong basis for cooperative effort.

---

[2] Max Radin, The Law and Mr. Smith (N.Y.: Bobbs-Merrill, 1938), p. 11.

As Professor Hans Zeisel has pointed out, the information in the social sciences which is relevant for the law may come from experiments, surveys, and so on. And he has made it clear that the validity and the usefulness of such information are directly tied to the methods by which it is obtained. Indeed, the portions of his discussion which concern the concepts of regression, random sampling, and variance are hardly understandable without some familiarity with methodological issues--and this is not merely a matter of jargon, special vocabularies, or statistics but involves the most basic notions of research.[3]

All this is to say that in joining law and the social sciences, a knowledge of methodology is essential. There is always the danger when the materials of one discipline are wedded to another that the limitations of the materials will be ignored or misunderstood--or that an excessive scepticism will come into play and valid findings will be rejected with the application of an unwarranted rigor. Thus, questions about the reliability of social science evidence, raised by Professor Edmond Cahn in discussing the U.S. Supreme Court's decision in the social segregation cases, are to be seriously considered. Professor Clark's rejoinder, however, vigorously argues that a criticism of such evidence must be based on more than a "personal, subjective reaction."

---

[3] See Hans Zeisel, "The Law," in: Lazarsfeld, Paul F., Sewell, William H., and Wilensky, Harold (Eds.), The Uses of Sociology (New York: Basic Books, Inc., 1967) pp. 81-91.

Many writers have pointed out, it is true, that when the social sciences have nothing to talk about, they talk about methods. Interminable discussions of methodology can be a way of avoiding difficult substantive issues; and the study of techniques, except for a few scholars who make it their specialty, is a means rather than an end. Yet if the study of methods sometimes appears somewhat tedious (a preparation for a trip, as more than one writer has indicated, which all too often remains preparation for a trip that fails to materialize), it is still necessary. In fact, the study of methods can be stimulating, for it requires imagination and creativity, just as the study of substantive social theory requires imagination and creativity.

It is our fundamental argument, then, that in the growing use of social science by the law, the law student and the lawyer must be able to weigh the findings of the social sciences with a critical eye. And this means that the law student and the lawyer need a knowledge of how the social sciences accumulate their evidence and build their theories.

III

A student in law school cannot be expected to become highly skilled in the social sciences. In fact, there is no reason why he should do so.

He can, however, acquire the knowledge he needs for an evaluation of research and its application in legal areas. And he can become skillful enough to carry out a research project of his own, which we think is one of the best ways to learn about research.

When social science materials appear in legal affairs, the lawyer must ask: What are the assumptions and theoretical underpinnings of the assertions drawn from the social sciences? Is the research design adequate to support the conclusions? What limitations must be placed on generalizing from the group studied to a larger population? If statistical methods are used, are they appropriate? Do they hide or reveal the individual facts and the relationships among facts? Do the research findings warrant a particular program of action? These are the kinds of questions the lawyer is increasingly called on to answer, whether he is dealing with evidence in a trial, a proposal for the extension of legal services for the poor, or a study of the administration of justice. Our purpose--and admittedly it is a modest one--is to provide law students with a background so that they can begin to answer such questions.

The materials in this collection are designed to cover two quarters, with the first quarter spent on the readings and exercises. The second quarter would best be devoted, in our opinion, to independent research projects undertaken by law students. If the law student is better equipped

for his professional work--as a practicing attorney, as a legal policy maker, or as a legal scholar--we will have accomplished our aim.

LAW AND SOCIAL SCIENCE:
A LOOK AT THE PRESENT STATE OF AFFAIRS

The following article by Professor Nagel discusses the present involvement of the law with the social sciences. He sets out three aspects of the legal profession (practicing lawyer, legal policy maker, and legal scholar) and lists the various uses which each has made of the social sciences. His citation of cases, treatises, journals, and monographs is wide-ranging and provides a useful bibliography for students of socio-legal research.

From Stuart Nagel, "Law and the Social Sciences: What Can Social Science Contribute?" American Bar Association Journal (Copyright 1965) pp. 356-358. Reprinted by permission.

Stuart S. Nagel, Law and the Social Sciences:
What Can Social Science Contribute?
51 American Bar Association Journal, 356(1965)

The Section of Bar Activities of the American Bar Association has established a Committee on Law and the Social Sciences, under the chairmanship of Richard F. C. Hayden, a judge for the Superior Court of Los Angeles.[1] It is my purpose to offer some tentative answers to the question, "What can social science contribute to law?" It seems logical to answer this question from the point of view of the three main types of personnel within the legal profession: the practicing lawyer, the legal policy maker and the legal scholar. The specific examples given in the footnotes represent a sampling of the relevant literature.[2]

The Practicing Lawyer

Social science methodology and substance can be useful to the practicing lawyer mainly by providing him with some materials that he might be able to introduce into evidence to win specific cases or points.[3] With regard to methodology, polling techniques may be valuable in resolving such issues as trademark infringement,[4] change of venue[5] or community standards of fitness for citizenship.[6] Systematic statistical analysis may be relevant to show that an alleged favoritism or discrimination is not readily attributable to chance.[7] Systematic observation or content analysis of documents may reveal patterns of behavior relevant to determining what constitutes reasonable care in a negligence case or common usage in a contract dispute.[8]

With regard to substance, psychological information may be crucial in civil or criminal insanity cases[9] or in any case involving the perceptive powers of witnesses.[10] Economic information may be crucial in antitrust, labor-management or commercial disputes.[11] Information gathered from historical[12] or political science[13] sources may be relevant to interpreting a statute or other legal document; it may also be relevant to a constitutional or administrative law issue. Sociological knowledge may be important in the sentencing aspects of a criminal case or the disposition of a family law case.[14]

Social science methodology and substance, in addition to being of value to the practicing lawyer with regard to the presentation of evidence in specific cases, can also help to improve his techniques as a lawyer in general.

As to trial techniques, social science methods are capable of testing a variety of hypotheses concerning the relative effectiveness of alternative techniques. Professors Kalven and Zeisel of the

University of Chicago jury project, for instance, have played tape recordings of trials to many actual juries in order to determine how the juries would react to slight variations in the trials. Among other things, they found that if the plaintiff's lawyer mentions that the defendant has insurance and the defendant's lawyer successfully objects, then the juries will tend to award higher damages to the plaintiff than if the defendant's lawyer did not object at all. They also found that working-class jurors in personal injury cases are more likely to find liability but are less likely to award higher damages than upper-class jurors--a bit of information helpful in picking juries after one has assessed the probability that liability will be found.[15]

Social science potentially can contribute to the skillful handling of clients in one's office. The Jury Verdict Research Corporation of Cleveland, Ohio, for instance, publishes a looseleaf service indicating the percentage of a wide variety of cases that have been decided in favor of the plaintiff and the average damages awarded. This kind of information, systematically gathered from across the country, can be helpful in advising a client on an out-of-court settlement.[16] A number of social psychologists have made systematic studies of elements affecting bargaining and negotiation. These would also be of value in developing out-of-court settlement techniques.[17]

Linguistic analysis, which is becoming an increasingly important part of social science, can be of value to lawyers interested in drafting legal instruments that are more meaningful and precise.[18] Content analysis by computer, which is being developed by social scientists, shows promise of becoming a useful aid in legal research.[19]

The Legal Policy Maker

By legal policy maker in this context is meant a legislator, appellate court judge or upper-level administrator. Social science can possibly contribute to improving both the procedure of the policy maker and the substantive materials available to him.

With regard to procedure, social science studies have tested systematically a number of proposals to reduce delay in the courts, such as splitting the liability and damages decision,[20] providing for nonjudicial auditors[21] and a variety of other techniques.[22] Other procedural reforms have also been subjected to social science analysis, such as providing neutral medical testimony,[23] appointing rather than electing the judiciary,[24] using blue-ribbon juries,[25] providing counsel to the indigent[26] and releasing indigent criminal defendants without bail pending trial.[27] The last study, for example, showed that a careful screening and notifying of defendants released without any bond produced a higher percentage of court appearances than the traditional bail bond system.

As to substantive developments in the law, social science has been helpful in many fields. A revealing study was made, for instance, on the effects of replacing the contributory negligence rule with the comparative negligence rule.[28] The effects of capital punishment have also been subjected to social science scrutiny,[29] as have the effects of income taxes on incentives to work[30] and the $1.00 minimum wage on the labor market.[31] Of course, the effects of segregation and desegregation have been studied extensively.[32] There has recently been a call for more social science research aimed at improving compliance with international law[33] and into the effects of prayers and Bible reading in the public schools.[34] At a more down-to-earth level is the perceptive questionnaire study which was made to determine the extent to which inheritance laws conform to the intent of people who die without wills.[35]

The Legal Scholar

Legal scholars in this context refers to law professors and to practicing lawyers and policy makers who are interested in the theoretical aspects of the legal process. Social science can probably make its greatest contribution to legal theory by investigating the causal forces behind judicial, legislative and administrative decision making and by probing the general effects of such decisions.

With regard to the causal forces behind the lawmaking and law-applying processes, the anthropologists[36] and historians[37] especially can broaden the legal scholar's perspective of the cultural roots beneath his legal field. The impact of public opinion on some aspects of the legal system also has been scrutinized systematically.[38] Likewise, the interactions between governmental bodies as a factor in determining decisional outcomes has been studied[39] as has the role of pressure groups[40] and political parties.[41] Witnesses,[42] lawyers[43] and litigants[44] have all been the subject of systematic social science study. Studies have also been made of the relation between the backgrounds and attitudes of judges[45] and legislators[46] and their decisional behavior.

As to the general impact of broad fields of law contrasted to specific laws, a number of studies have been and are under way. One sociologist has been studying the relation of legal impact to informal and formal controls[47] and to negative and positive appeals.[48] Other social scientists have contributed theory and data to the analysis of sanctions and legal compliance in general,[49] in public law[50] and in private law.[51] Using extensive interviewing techniques, one legal scholar analyzed the impact of contract law on business practice.[52] The role of factors such as the mass media, which facilitate or inhibit the impact of the law, has also been studied.[53]

Although this paper has only dealt with what social science can contribute to law, it should be noted that no social scientist can really understand the American or any other society unless he has a reasonable understanding of its legal system. Nevertheless, there is much that lawyers can learn from social scientists. What is probably needed is a greater awareness by lawyers of the research of interest to them that social scientists and lawyers oriented to social science are doing. It is hoped that the Section's new Committee on Law and the Social Sciences will help to fill that need and stimulate further research.

---

[1] "Activities of Sections", 50 A.B.A.J. 283 (1964).

[2] For further discussion of the relations between social science and law, see LAW AND SOCIOLOGY (Evan ed. 1962); Davis et al., SOCIETY AND THE LAW--NEW MEANINGS FOR AN OLD PROFESSION (1962); and Frontiers of Legal Research, 7 Am. Behav. Sci. 1 (December, 1963).

[3] Greenberg, Social Scientists Take the Stand--A Review and Appraisal of Their Testimony in Litigation, 54 Mich.L.Rev. 953 (1956).

[4] Zeisel, The Uniqueness of Survey Evidence, 45 Cornell L.Q. 322 (1960).

[5] Sherman, Use of Public Opinion Polls in Continuance and Venue Hearings, 50 A.B.A.J. 357 (1964).

[6] Repouille v. United States, 165 F. 2d (1947), especially Jerome Frank's dissent.

[7] Ulmer, Supreme Court Behavior in Racial Exclusion Cases, 56 Am. Pol.Sci.Rev. 325 (1962).

[8] Selltiz et al., RESEARCH METHODS IN SOCIAL RELATIONS 199-234 and 315-342 (1962).

[9] Lassen, The Psychologist as an Expert Witness in Assessing Mental Disease or Defect, 50 A.B.A.J. 239 (1964).

[10] McCarty, LEGAL AND CRIMINAL PSYCHOLOGY (1961).

[11] Economics in Antitrust Policy and Practice, 20 A.B.A. Antitrust Section (1962).

[12] Kelly, The Fourteenth Amendment Reconsidered, 54 Mich.L.Rev. 1049 (1959).

[13] Reynolds v. Sims, 377 U.S. 533 (1964), cites numerous political scientists on various aspects of the reapportionment controversies.

[14] Rose, The Social Scientist as an Expert Witness, 40 Minn.L.Rev. 205, 209 (1956).

[15] Broader, University of Chicago Jury Project, 38 Neb.L.Rev. 744 (1959).

[16] Nagel, Statistical Prediction of Verdicts and Awards, 63 M.U.L.L. 135 (1963).

[17] Sacks, Human Relations Training for Law Students and Lawyers, 11 J.Legal Ed. 316 (1959).

[18] Allen, Some Uses of Symbolic Logic in Law Practice, 8 Prac.Law 51 (1962).

[19] Gurr & Panofsky (eds.), Information Retrieval in the Social Sciences--Problems, Programs, and Proposals, 7 Am.Behav.Sci. 1 (June, 1964).

[20] Zeisel & Callahan, Split Trials and Time Saving--A Statistical Analysis, 76 Harv.L.Rev. 1606 (1963).

[21] Rosenberg, Auditors in Massachusetts as Antidotes for Delayed Civil Courts, 110 U.Pa.L.Rev. 27 (1961).

[22] Zeisel, Kalven & Bucholz, DELAY IN THE COURT (1959).

[23] Assoc. Bar N.Y.C., IMPARTIAL MEDICAL TESTIMONY (1956).

[24] Nagel, *Political Party Affiliation and Judicial Decisions*, 55 Am.Pol.Sci.Rev. 848 (1961).

[25] Fay v. New York, 332 U.S. 261 (1947).

[26] Silverstein, DEFENSE OF THE POOR IN CRIMINAL CASES IN AMERICAN STATE COURTS--A PRELIMINARY SUMMARY (1964). This study, made possible by the American Bar Foundation, was presented to the House of Delegates of the American Bar Association in August of 1964. The House of Delegates adopted recommendations for standards for representation of the indigent accused. 50 A.B.A.J. 969 (1964).

[27] Ares, Rankin & Sturz, *Manhattan Bail Project*, 38 N.Y.U. L.Rev. 67 (1963).

[28] Rosenberg, *Comparative Negligence in Arkansas*--*A Before and After Survey*, 13 Ark.L.Rev. 89 (1959).

[29] Sellin, THE DEATH PENALTY (1959) (a report for the Model Penal Code project).

[30] Break, *Income Taxes and Incentives To Work*--*An Empirical Study*, 47 Am.Econ.Rev. 529 (1957).

[31] Douty, *Some Effects of the $1.00 Minimum Wage in the U.S.*, 27 Economica 137 (1960).

[32] Myrdal, AN AMERICAN DILEMMA (1944).

[33] Larson, INTERNATIONAL RULE OF LAW (1961).

[34] Abington School District v. Schempp, 374 U.S. 203, 290, 319 (1963).

[35] Dunham, *Method, Process, and Frequency of Wealth Transmission at Death*, 30 U.Chi.L.Rev. 1 (1963).

[36] Hoebel, LAW OF PRIMITIVE MAN (1954).

[37] Kempin, LEGAL HISTORY--LAW AND SOCIAL CHANGE (1963).

[38] Rose and Prell, *Does the Punishment Fit the Crime*--*A Study in Social Valuation*, 61 Am.J.Soc. 247 (1955).

[39] Murphy, *Lower Court Checks on Supreme Court Power*, 53 Am.Pol. Sci.Rev. 1017 (1959).

[40] Vose, *Litigation as a Form of Pressure Group Activity*, 319 Annals 20 (1958).

[41] Sayre & Kaufman, GOVERNING NEW YORK (1960).

[42] Fishman & Morris (eds.), *Witnesses and Testimony at Trials and Hearings*, 13 J.Soc.Issues 1 (1957).

[43] Ladinsky, *Careers of Lawyers, Law Practice, and Legal Institutions*, 28 Am.Soc.Rev. 47 (1963).

[44] Hunting & Neuwirth, WHO SUES IN NEW YORK CITY (1962).

[45] Nagel, Judicial Backgrounds and Criminal Cases, 53 J.Crim.L., Crim. & P.S. 333 (1962).

[46] Wahlke et. al., THE LEGISLATIVE SYSTEM--EXPLORATIONS IN LEGISLATIVE BEHAVIOR (1962).

[47] Schwartz, *Social Factors in the Development of Legal Control*, 63 Yale L.J. 471 (1954).

[48] Schwartz, *Field Experimentation in Socio-legal Research*, 13 J.Legal Ed. 401 (1961).

[49] Arens & Lasswell, IN DEFENSE OF THE PUBLIC ORDER--THE EMERGING FIELD OF SANCTION LAW (1961).

[50] Ball, *Social Structure and Rent Control Violations*, 65 Am.J.Soc. 598 (1960).

[51] Litwak, *Three Ways in Which Law Acts as a Means of Social Control--Punishment, Therapy, and Education*--*Divorce as a Case in Point*, 34 Soc. Forces 217 (1956).

[52] Macaulay, *Non-Contractual Relations in Business*, 28 Am.Soc.Rev. 55 (1963).

[53] Newland, *Press Coverage of the U.S. Supreme Court*, 17 W.Pol.Q. 15 (1964)

THE SOCIAL SCIENTIST AND THE JUDICIAL PROCESS

Edmond Cahn, in criticizing Kenneth Clark's study of the negative effects of racially segregated schools, argues that the evidence presented hardly deserves to rank as scientific proof.

Professor Cahn raises a serious question concerning the status of social science findings when applied to everyday affairs. He asks, in effect, why should a court (or anyone else, for that matter) accept social science findings as being anything more than common sense observation? While we agree with the legitimacy of such a question, we feel that it is well answered by Professor Clark.

Professor Cahn's article expresses, we think, some indignation over the role to be played by social scientists in traditionally legal problems. And efforts such as those of Professor Clark and other behavioral scientists are likely to be seen as an encroachment on the advocate's role, a threat to the legal profession. Such a position is, we believe, quite unwarranted. How the findings of the social sciences are used is a question of legal strategy that must be decided and carried out by the lawyers. When there is a question concerning the weight to be ascribed to such empirical evidence, it becomes the task of the advocate to explain the methodology, the scientific basis of the research on which the evidence is founded.

From Edmond Cahn, "Jurisprudence," 30 New York University Law Review, (1955) pp. 150-169. Reprinted by permission.

From Kenneth B. Clark, "The Desegregation Cases: Criticism and the Social Scientist's Role," Villanova Law Review, Vol. 5, No. 2, pp. 224-235. Copyright 1959 by Villanova University. Reprinted by permission.

Edmond Cahn, Jurisprudence, 30 New York
University Law Review, 150 (1955)

In legal philosophy it is a very dangerous thing to comment on happenings that one welcomes and approves. By doing so, an unwary observer may expose himself to the occupational disease of jurisprudence, which in its most virulent form consists in finding the law "the true embodiment of everything that's excellent." When the disease takes hold, its victim gradually loses his early vigor of discontent and passion for reform; by slow stages he subsides into a state of complacency; in course of time, sitting at ease among his memories, he grows to resent the process of change as ardently as he used to welcome it. To him the crop of occurrences seems annually to present more and more of the irremediable, though not necessarily less of the evil, in juristic affairs; and so at last he becomes a recognized sage and elder statesman of the profession. The risk is a dire one. Nevertheless, in the teeth of the danger, I shall proceed to comment here on 1954's most important development: the Supreme Court's decisions in the School Segregation cases.[1] When the sense of injustice has triumphed on a scale so extensive, comprehension of the meaning of victory may prove useful to it in other conflicts. The decisions have added to the dignity and stature of every American.

The Opinions

Considerations of Style.--There is no need to speculate on the Supreme Court's unanimity in Brown v. Board of Education and Bolling v. Sharpe.[2] Before the opinions were announced, there may have been considerable room for doubt concerning their purport, but, at least in the eyes of sophisticated observers, there was hardly any question that the court would be unanimous or nearly unanimous. The train of previous decisions[3] dealing with the same subject matter had long since established this institutional policy.

Chief Justice Warren's style in the Brown and Bolling opinions is most commendably bland. Very wisely he declined the opportunity --one may say, the temptation--to indulge in democratic rhetoric. There is not a word of reproach or provocation in either opinion. To his credit, the Chief Justice was less concerned with getting into the anthologies than with presenting the country with a model of rational calm.

Two Canards.--Before going deeper into the cases, it is just as well to clear the air of some historical fictions concerning them. History is a grab-bag of such diverse and contradictory contents that a theorist is likely to find there whatever he needs for his immediate purpose. But to say that it will furnish evidence in support of almost any theory does not mean that the evidence it does furnish should be misrepresented or distorted.

On the one hand, therefore, the current practice of turning back to 1896 when Plessy v. Ferguson[4] upheld "separate but equal" facilities, and laying the blame for that decision, as some now do, at the door of a conciliatory Negro leader like Booker T. Washington, is anachronistic and unhistorical. Washington's policy may have been wrong; if he erred, his mistake consisted mainly in expecting the majority of Southern whites to become educated in the principles of racial equality faster than in fact they did. Condemning him at this late date seems uncharitable; it is always easier to be right sixty years after the event. Moreover, those who attribute the Plessy decision to Booker T. Washington's Atlanta speech of 1895 employ a very strange theory of social causation.[5] To say that the decision was the result of the 1895 speech, one must overlook everything else that had happened in the thirty years since Lincoln was assassinated. At worst, Washington's address may have been a straw in the fierce wind of bigotry.

The other canard is that the Brown and Bolling cases were decided as they were for the purpose of placating foreign countries whose good will had been jeopardized by racial segregation in the United States. Of course, it is possible that some few white people in the South will reconcile themselves more readily to mixed schools if they believe that their country's international position will be advanced by ending discrimination. But so many white Southerners are already willing to do away with segregated schools that considerations of the international order can hardly be given much weight. The same is true of the Supreme Court: while one trusts that the simplicity and clarity of the two opinions were intended to facilitate reading abroad as well as at home, it would be quite fallacious to suggest that the decisions themselves were motivated by subservience to foreign reactions. From Mr. Justice Harlan's ringing dissent in the Plessy case down through the sequence which began in 1938 with the Gaines decision,[6] there is demonstrative proof that the Supreme Court is perfectly capable of reading the Fourteenth Amendment through lenses molded and ground in the United States.

Edmond Cahn

Histories and the Constitution.--Among the many interesting aspects of these opinions, the Court's mode of interpreting the Fourteenth Amendment deserves very special attention. Never was Thomas Jefferson more clearly vindicated in his insistence that the Constitution belongs to the living generation of Americans. The question has become: What do the phrases mean here and now? The meaning of the Amendment is to be found in the relation of its language to a total haecceity, including everything that has changed and everything we have learned and unlearned since 1868 when it was adopted.[7] There were two lines of history for the Court to evaluate. The first of them--the history of the framing and adopting of the Fourteenth Amendment--proved to be only a casual element in the social and constitutional complex, an element far too ambiguous to be considered very important, much less decisive.

This was demonstrated rather adroitly. In an order for reargument,[8] the Justices requested the parties to dredge up whatever they could to show what the men of post-Civil War days had had in view concerning public education; zealously the various parties complied; after heavy labors, they adduced evidence as contradictory, amorphous, and confused as might have been--and, I suspect, was anticipated; whereupon the Court, with little show of disappointment, turned brusquely away from the mind of the nineteenth century and proceeded to put the mind of the twentieth century to work. The historical data continue to make interesting matter for reading, and not much more. Meanwhile, the order for reargument had hinted the outcome of the cases broadly enough to remove any resentment based on mere surprise.

Turning from history of the Amendment to history of the development of public education, the Court provided itself with much more consistent and objective indicia. One can see rather clearly the progression from what public education had been in 1868, when the Amendment was adopted, to 1896, when Plessy v. Ferguson was decided, and on to what it has become in our time. In view of this enormous growth, the Chief Justice rightly emphasized the new and paramount significance of free education in the political and cultural life of mid-twentieth century America. Surely then, the past cannot be allowed to decide for us what it did not have to decide for itself.

Some Dialectic Legerdemain.--The strategic considerations underlying the Brown and Bolling cases are fascinating to spell out. Apparently, the Justices determined that, in striking down segregated education, they were accomplishing quite enough for present purposes, and that the chances of obtaining cooperation from the rank and file of white Southerners would be reduced if the decisions should seem to touch even by implication on wider issues. If this is the way they reasoned, one must further commend their sagacity. On this reasoning, Plessy v. Ferguson must not be disturbed except as it related to matters of public education.

The 1896 case, it will be remembered, did not bear in any direct sense on issues of segregated education. Holding only that segregated transportation did not violate the Fourteenth Amendment, Plessy seemed easy to distinguish. But the problem was not so obligingly simple, for the Plessy opinion had not only recounted the early development of segregated public schools, but had reasoned more or less explicitly that segregated transportation was valid just because it followed the same pattern as segregated educational facilities. These passages were embarrassing. Chief Justice Warren had, in effect, to snatch the cloth of reasoning from under the Plessy case without spilling the content of its holding. If one examines the Brown opinion with these considerations in mind, the Chief Justice will appear to have executed his difficult assignment quite deftly.

The Brandeis-Brief Dilemma.--Professor Paul Freund has indicated more than once that while the Brandeis brief, filled with sociological and economic data for the judges' information, is an excellent device for upholding legislation, it creates an awkward logical predicament when the objective becomes one of overturning legislation.[9] If statistics, expert opinions, graphs, and similar data are sufficient to establish that the legislative or administrative authorities have acted rationally, then is it ever possible to prove that they have acted otherwise? Shrewd, resourceful lawyers can put a Brandeis brief together in support of almost any conceivable exercise of legislative judgment. Moreover, the mere fact that the legislature has acted in the premises ought normally to imply that the action had some rationally explicable basis.

The Brown and Bolling records, for example, included considerable evidence to show that school segregation was not unreasonable. In addition to explicit provisions in the respective state constitutions and statutes, there was the testimony of experienced school administrators, supported to some extent by the expert opinions of psychological and psychiatric witnesses (especially in the case from Virginia). Hence, whatever else the Supreme Court's action may accomplish, it does demonstrate satisfactorily that a challenge to constitutionality can overcome the kind of defense which is implicit in a Brandeis brief.

Edmond Cahn

It is fair to suspect that the impact of the Brandeis brief is no longer so great as when the device was novel and judges were more readily impressed by the paraphernalia of science or pseudoscience. In the last two decades, many Brandeis briefs have been conspicuously vulnerable in respect of statistical method, rationality of inferences from assembled data, adequacy of sampling, and failure to allow for--or to disclose--negative instances. Perhaps their quality will improve with a more critical attitude on the judges' part. . . .

Science or Common Sense?

A Dangerous Myth.--In the Virginia case[14] and to a lesser extent in the other litigations, various psychiatrists, psychologists, and social scientists gave expert testimony concerning the harmful effects of segregation on Negro school children. In addition, some of appellants' witnesses prepared an elaborate statement on the subject, which, signed by a total of thirty-two experts, was submitted to the Supreme Court as an appendix to appellants' brief.[15] In the months since the utterance of the Brown and Bolling opinions, the impression has grown that the outcome, either entirely or in major part, was caused by the testimony and opinions of the scientists, and a genuine danger has arisen that even lawyers and judges may begin to entertain this belief.[16] The word "danger" is used advisedly, because I would not have the constitutional rights of Negroes--or of other Americans--rest on any such flimsy foundation as some of the scientific demonstrations in these records.[17]

The moral factors involved in racial segregation are not new--like the science of social psychology--but exceedingly ancient. What, after all, is the most elementary and conspicuous fact about a primitive community if not the physical proximity of human beings mingling together? When the members of a community decide to exclude one of their number from the group life without killing him outright, what else can they do but force him to remove himself physically (as in the case of Cain), ostracize him for what they consider the general welfare (as the Athenians did), banish him from the cluster of community dwellings (as in outbreaks of leprosy or other plague), assign him a fixed area or ghetto to occupy (as with the Jews in medieval times), or lock him in a penitentiary (as we do with convicted criminals)? Hardly anyone has been hypocritical enough to contend that no stigma or loss of status attaches to these forms of physical separation. Segregation does involve stigma; the community knows it does. It knows full well that if "Stone walls do not a prison make nor iron bars a cage," they certainly do hamper a person's freedom to move about and consort with whom he pleases. Possibly, as the poet said, the walls can be understood as a "hermitage" or retreat or monastery, but only for those who choose them without being compelled by the social group.

There are people who argue, sometimes quite sincerely, that racial segregation is not intended to humiliate or stigmatize. On first impression, the argument seems to have some slight mitigative value, for surely a deliberate insult is liable to cut deeper than one inflicted out of mere crudeness or insensibility. But the mitigation comes too late. An excuse that one did not intend to injure does not stand much chance of reception when the offender, having been informed of the damage he has done, continues and persists in the same old callous insults. As is observed in the ancient Babylonian Talmud, to shame and degrade a fellow-creature is to commit a kind of psychic mayhem upon him. Like an assailant's knife, humiliation slashes his self-respect and human dignity. He grows pale, the blood rushes from his face just as though it had been shed.[18] That is why we are accustomed to say he feels "wounded."

Moreover, if affronts are repeated often enough, they may ultimately injure the victim's backbone. We hear there are American Negroes who protest they do not feel insulted by racially segregated public schools. If there are any such Negroes, then they are the ones who have been injured most grievously of all, because segregation has shattered their spines and deprived them of self-respect.

So one speaks in terms of the most familiar and universally accepted standards of right and wrong when one remarks (1) that racial segregation under government auspices inevitably inflicts humiliation, and (2) that official humiliation of innocent, law-abiding citizens is psychologically injurious and morally evil. Mr. Justice Harlan and many other Americans with responsive consciences recognized these simple, elementary propositions before, during, and after the rise of "separate but equal." For at least twenty years, hardly any cultivated person has questioned that segregation is cruel to Negro school children. The cruelty is obvious and evident. Fortunately, it is so very obvious that the Justices of the Supreme Court could see it and act on it even after reading the labored attempts by plaintiffs' experts to demonstrate it "scientifically."

Claims and Facts.--When scientists set out to prove a fact that most of mankind already acknowledges, they may provide a rather bizarre spectacle. Fifty years ago, certain biologists who were engaged in just this sort of enterprise, provoked George Bernard Shaw to denounce their "solemnly offering us as epoch-making discoveries their demonstrations that dogs get weaker and die if you give them no food; that intense pain makes mice sweat; and that if you cut off a dog's leg the three-legged dog will have a four-legged puppy."[19] Then Mr. Shaw called the scientists a number of fearful names (beginning with "dolts" and "blackguards"), none of which would be remotely applicable to the psychologists and psychiatrists who testified in the desegregation cases. So far as I can judge, all of these are fine, intelligent, dedicated scholars. Yet one can honor them as they deserve without swallowing their claims.

Professor Kenneth B. Clark of the psychology department of City College acted as general social science consultant to the NAACP legal staff and served as liaison between the lawyers and the scientists. His endeavors having been long and arduous, perhaps it was natural that he should exaggerate whatever the experts contributed to the case. In an article written while the country was waiting for the Supreme Court's decisions, he asserted, "Proof of the arguments that segregation itself is inequality and that state imposed racial segregation inflicts injuries upon the Negro had to come from the social psychologists and other social scientists."[20] (Emphasis supplied.)

When Professor Clark wrote thus, he could not know that Chief Justice Warren's opinions would not mention either the testimony of the expert witnesses or the submitted statement of the thirty-two scientists. The Chief Justice cushioned the blow to some extent by citing certain professional publications of the psychological experts in a footnote, alluding to them graciously as "modern authority."[21] In view of their devoted efforts to defeat segregation, this was the kind of gesture a magnanimous judge would feel impelled to make, and we are bound to take satisfaction in the accolade. Yet, once the courtesy had been paid, the Court was not disposed in the least to go farther or base its determination on the expert testimony.

As I have said, these developments Professor Clark could not have known when he staked so wide a claim for his profession. But he did know that circumstances in the Virginia litigation--the one he participated in most actively--had reflected very directly on his assertion. The Virginia school board had offered the testimony of three expert witnesses, and all three (psychiatrist Kelly, psychologist Buck, and Professor Clark's own former teacher, Professor Garrett of Columbia) had admitted, in one way or another, that racial segregation in the schools does injure Negro children's personalities.[22] They admitted, as we have said, a fact of common experience. On the defendants' behalf, they testified as persuasively as they could against the Court's adopting what they called "disruptive" or "coercive" measures, and they spoke regretfully about the firmness of established regional customs. Buck summarized their attitude in two statements: "I feel that as an abstract idea, segregation is bad," and "I think the whole society is sick."[23]

When we come to explain why the statement signed by the thirty-two social scientists went without mention by Chief Justice Warren, I find myself at a disadvantage. Only the reader's assistance can rescue me. I have examined the text of this statement, which has become easy of access by being reprinted in a law review.[24] My personal, subjective reaction is that the text conveys little or no information beyond what is already known in "literary psychology" (by which I mean such psychological observations and insights as one finds continually in the works of poets, novelists, essayists, journalists, and religious prophets). The statement's vocabulary and style would not be called "literary"; I refer only to its substance. If my readers will inspect the statement for themselves, they will ascertain whether it impresses them as it did me. At that, my reaction may be due to a lack of technical training in scientific psychology and psychological testing.

The "Generally Accepted" Test.--When a scientist is engaged in demonstrating a fact of common knowledge (e.g., that fire burns, that a cold causes snuffles, or that segregation degrades), it is not easy to pass a fair judgment on the validity of his proof. Our minds tend to supply his conclusion before he is ready to deduce it. Subconsciously we reinforce his evidence with the facts and feelings of our own experience, and if his reasoning should contain a flaw, we are too preoccupied with reaching the familiar destination to detect it. Moreover, in the present situation, men who specialize in conducting psychological tests might discover all sorts of weak assumptions and fallacies that mere lawyers would never notice. Under these several disadvantages, the most I can do here is present Professor Clark's evidence concerning the "generally accepted" test,[25] together with the comments that suggest themselves to an untrained but interested observer.

Professor Clark testified as an expert in the South Carolina, Delaware, and Virginia litigations. The clearest description of the test appears in his testimony in the South Carolina case. He said:[26]

> A. I made these tests on Thursday and Friday of this past week at your request, and I presented it to children in the Scott's Branch Elementary school, concentrating particularly on the elementary group. I used these methods which I told you about--the Negro and White dolls--which were identical in every respect save skin color. And, I presented them with a sheet of paper on which there were these drawings of dolls, and I asked them to show me the doll--May I read from these notes?
>
> Judge Waring: You may refresh your recollection.
>
> The Witness: Thank you. I presented these dolls to them and I asked them the following questions in the following order: "Show me the doll that you like best or that you'd like to play with," "Show me the doll that is the 'nice' doll," "Show me the doll that looks 'bad'," and then the following questions also: "Give me the doll that looks like a white child," "Give me the doll that looks like a colored child," "Give me the doll that looks like a Negro child," and "Give me the doll that looks like you."
>
> By Mr. Carter:
>
> Q. "like you?"
>
> A. "Like you." That was the final question, and you can see why. I wanted to get the child's free expression of his opinions and feelings before I had him identified with one of these two dolls. I found that of the children between the ages of six and nine whom I tested, which were a total of sixteen in number, that ten of those children chose the white doll as their preference; the doll which they liked best. Ten of them also considered the white doll a "Nice" doll. And, I think you have to keep in mind that these two dolls are absolutely identical in every respect except skin color. Eleven of these sixteen children chose the brown doll as the doll which looked "bad." This is consistent with previous results which we have obtained

testing over three hundred children,[27] and we interpret it to mean that the Negro child accepts as early as six, seven or eight the negative stereotypes about his own group. And, this result was confirmed in Clarendon County where we found eleven out of sixteen children picking the brown doll as looking "bad," when we also must take into account that over half of these children, in spite of their own feelings,--negative feelings--about the brown doll, were eventually required on the last question to identify themselves with this doll which they considered as being undesirable or negative. It may also interest you to know that only one of these children, between six and nine, dared to choose the white doll as looking bad. The difference between eleven and sixteen was in terms of children who refused to make any choice at all and the children were always free not to make a choice. They were not forced to make a choice. These choices represent the children's spontaneous and free reactions to this experimental situation. Nine of these sixteen children considered the white doll as having the qualities of a nice doll. To show you that that was not due to some artificial or accidental set of circumstances, the following results are important. Every single child, when asked to pick the doll that looked like the white child, made the correct choice. All sixteen of the sixteen picked that doll. Every single child, when asked to pick the doll that was like the colored child; every one of them picked the brown doll. My opinion is that a fundamental effect of segregation is basic confusion in the individuals and their concepts about themselves conflicting in their self images. That seemed to be supported by the results of these sixteen children, all of them knowing which of those dolls was white and which one was brown. Seven of them, when asked to pick the doll that was like themselves; seven of them picked the white doll. This must be seen as a concrete illustration of the degree to which the pleasures[28] which these children sensed against being brown forced them to evade reality--to escape the reality which seems too overburdening or too threatening to them. This is clearly illustrated by a number of these youngsters who, when asked to color themselves--For example, I had a young girl, a dark brown child of seven, who was so dark brown that she was almost black. When she was asked to color herself, she was one of the few children who picked a flesh color, pink, to color herself. When asked to color a little boy, the color she liked little boys to be, she looked all around the twenty-four crayons and picked up a white crayon and looked up at me with a shy smile and began to color. She said, "Well, this doesn't show." So, she pressed a little harder and began to color in order to get the white crayon to show. These are the kinds of results which I obtained in Clarendon County.

Q. Well, as a result of your tests, what conclusions have you reached, Mr. Clark, with respect to the infant plaintiffs involved in this case?

A. The conclusion which I was forced to reach was that these children in Clarendon County, like other human beings who are subjected to an obviously inferior status in the society in which they live, have been definitely harmed in the development of their personalities; that the signs of instability in their personalities are clear, and I think that every psychologist would accept and interpret these signs as such.

Q. Is that the type of injury which in your opinion would be enduring or lasting?

A. I think it is the kind of injury which would be as enduring or lasting as the situation endured, changing only in its form and in the way it manifests itself.

Mr. Carter: Thank you. Your witness.

General Comments.--We are not provided here with any proof of the numerical adequacy of the sampling or of its being a representative cross-section. We have no demonstration that abnormal or eccentric backgrounds of the individual children have been investigated. Among these 16 children (or 300, including the other groups mentioned) there would probably be a certain proportion with untypical private experiences. In such a strikingly small sample, the results could easily mislead.

Moreover, if one follows the arithmetic in Professor Clark's testimony--which is not easy for me--some of his interpretations seem to be predetermined. For example, if Negro children say a brown doll is like themselves, he infers that segregation has made them conscious of race; yet if they say a white doll is like themselves, he infers that segregation has forced them to evade reality.

Perhaps the main point is that this test does not purport to demonstrate the effects of school segregation, which is what the court was being asked to enjoin. If it disclosed anything about the effects of segregation on the children, their experiences at school were not differentiated from other causes. Considering the ages of the children, we may conjecture they had not been long at school.

Comment on the Opening Questions.--We do not know how the children took these questions. If Professor Clark had offered to give real dolls instead of showing pictures of dolls, the reaction might have been more serious. In any case, I do not think any certain inference follows from 10 out of 16 pointing to the picture of the white doll. Habituation with dolls (as distinguished from people) should be allowed for. Manufacturers and commercial fashions practically restrict a child's concept of what a "nice" doll would look like. Many white children of certain generations were taught to prefer "Topsy" or other colored dolls;[29] some children would say that there is no really "nice" doll but a teddy-bear. At this point, the response seems uninformative.

Comment on the "Bad Doll" Question.--Here, it seems to me, the children were tricked. Perhaps that is how some of them felt. There had been no previous question about a "good doll," only about a "nice" one, which the children clearly understood meant one "you'd like to play with." What is a "bad doll"? Some children might consider this a term of preference for play purposes: all little "mothers" love to rebuke and punish naughty dolls. Other children, on hearing the question, would be simply bewildered by the sudden, unexpected introduction of moral or disciplinary references. Some may have responded by pointing to the brown doll because the question seemed to imply that a process of elimination was contemplated. But I hope the children asked themselves: Why must there be a "bad" doll at all? Why cannot both dolls be "nice"? We observe that five children declined to answer this question. Probably they felt it unfair or at least very confusing in the circumstances.

Comment on the Remaining Questions.--It is noteworthy that seven Negro children picked the white doll "when asked to pick the doll that was like themselves." Professor Clark leaps to infer that they were evading reality. This I doubt. Although his testimony does not make me clear on the point, I gather that these seven children were among the ten who had previously chosen the white doll as "nice." Were they wrong, then, to claim that the white doll was very much "like themselves" because they too were "nice"? No one can state positively what these children were thinking at the time; but if they did have perception enough to insist to themselves that the "niceness" was decisive and not the color, lo and behold! this would be wisdom indeed! "Out of the mouths of babes and sucklings--"? Perhaps, merely perhaps. In any event, I cannot see that the opposite interpretation (Professor Clark's) is so evident that it deserves to rank as scientific proof.

Aid from an Unexpected Quarter.--Fortunately, the outcome of the *Brown* and *Bolling* cases did not depend on the psychological experts' facing and answering the objections, queries, and doubts I have presented. It is possible that if the questions had been put to Professor Clark on cross-examination, he would have come forward with convincing answers. But, to all intents and purposes, the questions were not put. The doll test was not analyzed in suitable detail by any of the cross-examiners, probably because they, too, realized that segregation does degrade and injure Negro school children.

In the Virginia trial, the defense appeared particularly inept. Far from caring to concentrate on the doll test and its scientific validity, the lawyer for the defendants was preoccupied with other lines of cross-examination. He had a different set of values to display. Why concern himself with dissecting the experts' logic and the correctness of their inferences? Instead, questions were asked which would convey disparaging insinuations about a professor's parents, his ancestral religion, the source of his surname, the pigmentation of his skin, or the place of his birth. If these items did not discredit him satisfactorily, then one went on to inquire how many years he had spent in the South; if he had lived in the South, how long in Virginia; and so on--implying all the while that science, common sense, and human nature would not dare to cross Virginia county lines.[30] And, of course, there would be continual hints that what the plaintiffs' witnesses really desired to achieve was miscegenation and a mixed race.

As any healthy-minded person reads the Virginia trial record, it is impossible not to contrast the altruism and sober dignity of the scientists with the behavior of defendants' counsel, who, by his manner of espousing the old order, exposed its cruelty and bigotry.

Here was a living spectacle of what racial segregation can do to the human spirit. The segregated society, as defendants' own expert had said, was "sick"; and the tactics of cross-examination used by defendants' lawyer showed how very sick it was. I suggest that these pages of the record did not fail of notice in the deliberations of the United States Supreme Court.

Without Salt, No Science.--We may as well resign ourselves to letting the troglodytes remain troglodytes, and turn our attention back to our civilized friends, the social psychologists. As the courts' exclusionary rules of evidence tend to relax more and more, the scientists will appear more frequently to testify as expert witnesses. How much respect should the judges extend to their testimony?

The answer depends in large measure on the scientists. If I have been right in suggesting that their evidence in the desegregation cases seemed persuasive because it happened to coincide with facts of common knowledge, they surely cannot rely on having the same advantage in every future litigation. It is predictable that lawyers and scientists retained by adversary parties will endeavor more aggressively to puncture any vulnerable or extravagant claims. Judges may learn to notice where objective science ends and advocacy begins.[31] At present, it is still possible for the social psychologist to "hoodwink a judge who is not over wise" without intending to do so; but successes of this kind are too costly for science to desire them.

For one thing: Merely translating a proposition of "literary" psychology into the terms of technical jargon can scarcely make it a scientific finding. For another: Just because social psychology is in a youthful and somewhat uncertain stage, the utmost rigor should be imposed on its intermediate processes.

The point is vital, involving as it does not only social psychology's prestige in the courts but--what is ultimately more valuable--its capacity to evolve and progress as a cumulative body of tested knowledge and approved method. Among the major impediments continually confronting this science are (1) the recurrent lack of agreement on substantive premises, and (2) the recurrent lack of extrinsic, empirical means for checking and verifying inferred results. As long as these disadvantages remain, and they are likely to remain in some measure for a very long time, social psychology will need, above all things, the use of scrupulous logic in its internal, intermediate processes. If the premises must be loose, the reasoning from them should be so much tighter; and if the final results cannot be validated precisely by external tests, then the methods of inference should be examined and re-examined all the more critically. It is meticulous standards that bring respect and credence to scientific testimony. When a social psychologist is called to serve as a "friend of the court," he should be able to assume our belief that his best friend, his premier loyalty, is always the objective truth.

Some of the Consequences.--Obviously, the Brown and Bolling opinions are susceptible of more than one interpretation. My views do not agree with those of some very able commentators, who consider that the opinions show important marks of the psychologists' influence. Granting this variety of interpretations, does it really matter whether the Supreme Court relies or does not rely on the psychologists' findings? Does it make any practical different?

I submit it does. In the first place, since the behavioral sciences are so very young, imprecise, and changeful, their findings have an uncertain expectancy of life. Today's sanguine asseveration may be cancelled by tomorrow's new revelation--or new technical fad. It is one thing to use the current scientific findings, however ephemeral they may be, in order to ascertain whether the legislature has acted reasonably in adopting some scheme of social or economic regulation; deference here is shown not so much to the findings as to the legislature. It would be quite another thing to have our fundamental rights rise, fall, or change along with the latest fashions of psychological literature. Today the social psychologists--at least the leaders of the discipline--are liberal and egalitarian in basic approach. Suppose, a generation hence, some of their successors were to revert to the ethnic mysticism of the very recent past; suppose they were to present us with a collection of racist notions and label them "science." What then would be the state of our constitutional rights? Recognizing as we do how sagacious Mr. Justice Holmes was to insist that the Constitution be not tied to the wheels of any economic system whatsoever, we ought to keep it similarly uncommitted in relation to the other social sciences.

There is another potential danger here. It concerns the guarantee of "equal protection of the laws." Heretofore, no government official has contended that he could deny equal protection with impunity unless the complaining parties offered competent proof that they would sustain or had sustained some permanent (psychological or other kind of) damage. The right to equal protection has not been subjected to any such proviso. Under my reading of the Brown and Bolling opinions, this would remain the law. But if, in future "equal protection" cases, the Court were to hold that it was the expert testimony that determined the outcome of Brown and Bolling, the scope of the constitutional safeguard might be seriously restricted.[32] Without cataloguing the various possibilities, one can discern at least that some of them would be ominous. It is not too soon to say so, for basic rights need early alarms.

Whose Victory, Then?

When the American people suffered a moral defeat in Plessy v. Ferguson, Mr. Justice Harlan's dissent survived on the record. It reminded the people of the solemn promises and mutual pledges that had inspired their beginnings as a nation, and appealed from what was narrow and mean in their traditions to what was broad and fraternal. As years went by, Harlan's prophetic phrases were taken up, repeated, and reiterated--until they passed gradually into the popular vocabulary. Clergymen, social scientists, publicists, and jurists imagined they were coining novel phrases as they unwittingly borrowed his, which he would have been glad to lend them. By the end of the 1930s, when the Supreme Court began to utter its series of lessons to the states on fair treatment in education, the American people were fully prepared to support their judicial pedagogue. Now the advance has reached its climax, with the principle of equality firmly established. In this struggle, there can be no one to hail as victor except the whole American people, who alone have the power to insure that the principle becomes a living reality in their public schools. The achievement is theirs. With this, I believe, the lone dissenter of 1896 would agree. Even as the shame of his time belonged to all, so should the vindication of our time; and so too the hope of a more benevolent American community.

---

[1] Brown v. Board of Educ., 347 U.S. 483 (1954); Bolling v. Sharpe, 347 U.S. 497 (1954). The commentary will be limited to developments which reached completion in 1954.

An appendix to this article lists several items of value touching on other aspects of jurisprudence.

[2] Supra note 1. See the valuable symposium, Segregation in the Public Schools, 3 J. Pub. L. 5 (1954).

[3] Listed at 347 U.S. 492 (1954).

[4] 163 U.S. 537 (1896). There have been signs of a campaign to transform Mr. Plessy into a heroic asserter of Negro rights. This will take some doing, at least if people read Plessy's petition. He alleged that he was "seven eighths Caucasian and one eighth African blood; that the mixture of colored blood was not discernible in him, and that he was entitled to every recognition, right, privilege and immunity secured to the citizens of the United States of the white race," etc. Id. at 538.

[5] The Negro in American Life and Thought: The Nadir 1877-1901, pp. 278 et seq. (1954), a very useful survey by Professor Rayford W. Logan of Howard University. While criticizing Brooker T. Washington sharply, Professor Logan concedes: "Washington was convinced, and rightly so, that it would have been folly to ask in 1895 for equal rights for Negroes." Id. at 280.

[6] Missouri ex rel. Gaines v. Canada, 305 U.S. 337 (1938).

[7] This theme was developed most effectually by my colleagues who participated in the symposium, Supreme Court and Supreme Law (Cahn ed. 1954).

[8] 345 U.S. 972 (1952). For other tactical developments, see McKay, Segregation and Public Recreation, 40 Va.L.Rev. 697 (1954).

[9] Freund, "Review of Facts in Constitutional Cases" in Supreme Court and Supreme Law 49-50 (Cahn ed. 1954). For another analysis by the same gifted author, see On Understanding the Supreme Court 88 et seq. (1951).

[14] Davis v. County School Board, 103 F. Supp. 337 (E.D. Va. 1952), No. 4 in the Supreme Court.

[15] Reprinted, The Effects of Segregation and the Consequences of Desegregation: A Social Science Statement, 37 Minn.L.Rev. 427 (1953).

[16] For example, Will Maslow, Director of the Commission on Law and Social Action of the American Jewish Congress, said: "The NAACP struggled for sixteen years to convince the United States Supreme Court that segregation in and of itself was a form of inequality forbidden by the Fourteenth Amendment. As long as the argument revolved about the size of classes, the length of school terms, the salaries of teachers, the physical condition of the school plant or the distance required to travel to school, little headway was made in convincing the court. But when the psychologists began to argue about the injury to

pupil morale caused by governmentally imposed segregation, the sense of inferiority occasioned thereby and the hampering of the learning process, the Court pricked up its ears. When the final decision was handed down in the public school segregation cases, it rested not on conceptual legal principles or the legislative history of the Fourteenth Amendment or even on the sociological demonstration that in practice segregation results in inferior schools but on the psychological finding of thwarted intellectual development." Address, The Uses of Law in the Struggle for Equality, Atlantic City, N.J., December, 1954.

[17] Clearly, counsel followed the right strategy in offering the expert testimony for whatever the several courts or judges might find it worth. The younger Pliny, an extraordinarily successful advocate, insisted at some length that since different minds may be persuaded by different arguments, the advocate ought to develop and present them all, neglecting none. Pliny, Letters I. xx (Loeb ed. 1931) I, 70-2. (It is barely possible that a phrase in this passage suggested Holmes' metaphore of "the jugular.").

[18] Baba Mezia 58b in The Babylonian Talmud, Seder Nezikin II 348-49 (Soncino trans. 1935).

[19] From a lecture prepared by Shaw in 1906 for delivery before the Fabian Society. This is one of several excerpts which he used later in the Preface to Back to Methuselah.

[20] Clark, Desegregation: An Appraisal of the Evidence, 9 J. Social Issues No. 4, p. 3 (1953). It would relieve me to learn that I have read Professor Clark too literally, but general opinion indicates that I have not.

[21] 347 U.S. 483, 494 and n.11 (1954).

[22] No. 4, Davis v. County School Board (the Va. Case), Transcript of Record: Kelly testifying at p. 529; Buck at pp. 538, 544; and Garrett at p. 550. Professor Garrett said: "I do not think that one can possibly defend separation of one group from another, if the separated group is stigmatized or put into an inferior position. Separation can be of different sorts which does not involve, necessarily, any feeling of inferiority or any stigma." And then he went on to instance, as examples of separation without stigma, the following:

(1) separate schools for boys and girls, (2) parochial schools, (3) classes for "children who are slow," and (4) classes for "children who are bright." (May all the opposers of equality have witnesses like these!)

[23] Id. at 538, 544.

[24] See note 15 supra.

[25] For Professor Clark's testimony that this is the test "generally accepted as indications of the child's sensitivity to race as a problem," see No. 2, Briggs v. Elliott (the S.C. case), Transcript of Record, p. 86. Professor Clark suffered from inflicting the test on the Negro children. No. 4, Davis v. County School Board (the Va. case), Transcript of Record, p. 251.

[26] No. 2, Briggs v. Elliott (the S.C. case), Transcript of Record, pp. 87-90.

[27] A bit later, the number was stated as "400 times." Id. at 96.

[28] Thus in original; probably should be "pressures."

[29] Professor Clark testified that a graduate student at Columbia had used the test on white children with his permission, but he had not obtained the results. No. 2, Briggs v. Elliott (the S.C. case), Transcript of Record, p. 96. Yet it would seem that trying the test on white children would be the very first and most obvious way to begin ascertaining whether it had any probative value when given to Negro children.

[30] No. 4, Davis v. County School Board (the Va. case), Transcript of Record, pp. 214, 262, 265. By way of contrast, in 1954 the Virginia legislature enacted a statute prohibiting "any commercial advertisement that any person not otherwise prohibited by law from using an establishment is not welcome, or is objectionable, or is not acceptable because of his religion." Va. Code Ann. § 57-2.1 (Supp. 1954).

[31] In the attractive account cited in note 20 supra, Professor Clark described the close collaboration between the NAACP lawyers and the scientists, and went on to remark: "In fact, there were times when the lawyers could speak as social psychologists and the social psychologists began to sound like lawyers. In spite of this mutual accommodation, however, a clear distinction of roles and responsibilities had to be maintained for effective collaboration." Clark, supra note 20, at 6. It seems possible that the distinction of roles would be maintained more satisfactorily if the social psychologist's primary motive in maintaining it were strict fidelity to the objective truth rather than "effective collaboration."

For an example of the kind of methodological criticism we need from the social scientists, see Hoebel, The Law of Primitive Man 272-74 (1954). The instance discussed by Professor Hoebel likewise involved a very flabby attempt to "demonstrate" an assertion that on its face would seem entirely plausible to the lay mind.

Under the circumstances, some readers may feel inclined to credit Chief Justice Warren with Mephistophelean wit in that famous "footnote 11" (the accolade to the social psychologists which we have already mentioned). In the footnote, the Chief Justice lists various works published between the years 1944 and 1952. The latest of these discloses: "Unfortunately for scientific accuracy and adequacy, thoroughly satisfactory methods of determining the effects of prejudice and discrimination on health of personality have not yet been devised, nor has a sufficient number of studies dealing with the various minority groups been made." (Emphasis supplied.) Witmer and Kotinsky, Personality in the Making 139-40 (1952).

[32] In 1953 Annual Surv. Am. L. 811 n.30, 29 N.Y.U.L. Rev. 485 n.30 (1954, I proposed a canon for interpreting a certain limited class of writing. If valid, the canon would elucidate a writing in which, after a long controversy, one of the parties concedes that the opponent's position has been the right one. Without quoting or repeating last year's exposition, to which I hope some of my present readers will refer, I may summarize the "canon of concession" in three aspects: (1) viewed as rhetoric, the conceding disputant's statement will tend to emphasize any newly advanced reasons for coming to the conclusion he had previously rejected, because the new reasons are the face-savers; (2) viewed as definition, his statement will seem to imply that his adhesion to the conclusion is restricted by the new, face-saving reasons; and (3) viewed as prediction of his future behavior, the face-saving

reasons and implied restrictions tend gradually to disappear--unless some later occurrence should happen to renew their pertinence and force.

If I may venture now to apply this canon to the Brown opinion, treating it as a concession in the long debate with Mr. Justice Harlan's Plessy dissent, then it seems to suggest an adequate explanation of Chief Justice Warren's references to psychological writings. In this analysis, the data of social psychology would constitute, at most, convenient face-savers. Applied to them, the canon of concession helps us to understand their rhetorical usefulness. If they did have a definitional function (which I contest), it would also clarify that.

Is anyone temerarious enough to predict whether the Supreme Court will make further face-saving references to the data of social psychology in cases of this kind? Not I. However, it is fair to comment that, according to the canon of concession, the significance of any such references would sooner or later shrink to the disappearing point in "equal protection" doctrine--unless some independent factor should emerge to vivify them. Face-saving reasons generally do not outlast the occasion that calls them forth. So much may be said safely. Saying more would be wrapping the future in a formula, a thing foolish to attempt and impossible to accomplish.

*Kenneth B. Clark, The Desegregation Cases:
Criticism of the Social Scientist's Role
Willanova Law Review, Vol. 5 (1959)

Basic to the direct and indirect criticisms which have been raised concerning the role of social scientists in the school desegregation cases is the generally unstated question of the propriety of social scientists playing any role in this type of legal controversy. It is clear that the public school desegregation cases are crucially related to the delicate and specific problems of the relative status of the Negro and white groups in American culture and the equally delicate and general problem of social change. Before one attempts to discuss the specific criticisms or the fundamental questions which they appear to reflect, it might be valuable to attempt an analysis of the social dynamics, the context within which such discussions seem either necessary or desirable. Serious discussion of whether social scientists should play a role in the legal processes related to the desegregation of the public schools would seem no more or less justified than discussions of the following questions:

> Should social scientists play a role in helping industry function more efficiently--make larger profits--develop better labor management relations--increase the sense of satisfaction among the workers?
>
> Should social scientists play a role in helping governmental agencies and key policy makers make more effective and valid decisions?
>
> Should social scientists play a role in attempting to solve the many human and psychological problems faced by the military arm of our government?

The psychological significance of the fundamental problem posed by questioning the relationship between social scientists and the desegregation cases may be even more clearly illustrated by asking the analogous question:

> Should biological scientists play a role in guiding medical research and practices?

The answers to the above questions would seem so obviously positive that one is forced to question the validity of the question which is implicit in the criticisms which have been raised concerning the role of the social scientists in the desegregation cases. In searching for an answer, one must look in the direction of understanding the complexity of the power structure of our society and particularly the types of of threats to the existing social structure which are inherent in the recent decisions of the United States Supreme Court which ruled that racial segregation in public schools and other forms of state-supported public accommodations violate the equal protection clause of the fourteenth amendment of the United States Constitution. These decisions must be seen as demanding fundamental changes in the power alignments and group status patterns which prevail in our society. The social scientists who collaborated with the lawyers who argued and won these cases were certainly accessories to this demand for a significant form of social change. They themselves might not have been psychologically prepared to accept with equanimity the directness of the involvement or the sweeping demand for social change which the Court's decision precipitated. It is also possible that these changes are not only contrary to the prevailing status hierarchy among the racial groups in our society but also inimical to an important aspect of the continued controlling power pattern of this society. If this is true, the accessory role of social scientists in these decisions subjects them to the criticisms of those who are identified with and seek to perpetuate the racial status quo and the related power controls.

It may be, therefore, that the continued preoccupation of social scientists and their critics with the question of whether they should be involved in this phase of the legal processes reflect their anxiety in the face of these criticisms; and reflects even more concretely the possibility that these criticisms may lead to more punitive controls of those social scientists who continue to identify themselves with "controversial causes"--i.e., causes which threaten the prevailing power alignments in the society.

Social scientists, like other knowledgeable individuals in our society, must be sensitive to the problems of power and the techniques of social control which are operative in the society in which they work. In spite of the demand for objectivity and integrity in the search for truth, the important determinant of serious scientific work, social scientists are influenced indirectly and sometimes directly, subtly and sometimes crudely, by the prevailing social biases and uncritically accepted frames of reference of their society.

Given this perspective, one can then begin to evaluate the specific criticisms which have been raised against the social scientists who have been involved in these desegregation cases. The implications of any of these criticisms are not restricted to the more academic problems of social science theory, methodology, and the nature of social science evidence. Nor are they limited to the more complex problems of the delicate relationship between the social sciences and the law. These are indeed crucial problems which merit continuous discussion and debate in the relatively young and dynamic social sciences. The full import of a given criticism must be understood in terms of whether it clarifies or distorts the larger social issues; specifically, the practical reality of the nature, function, and consequences of racial segregation in American life, the stresses and strains inevitably involved in attempts to change institutionalized patterns of social injustices, and the role of the courts and other governmental agencies in the competition among groups for changes in, or maintenance of, the status quo.

Some of the most intense criticisms have come from political leaders of the deep southern states. Men like Senators Eastland and Talmadge, former Governor Byrnes and Governor Faubus have attacked the Supreme Court's decision not only on the grounds that it violated "states rights" but also, significantly for the purposes of this paper, on the grounds that it attempted to substitute psychological and sociological theories for the law. There is a question whether these types of criticisms should be taken seriously by social scientists since they seem motivated largely by political considerations.

Attacks on the role of social scientists in these cases have not been restricted to politicians who object to the Court's decision and the social changes which they fear may result, but have come also from serious students of jurisprudence and more recently from social scientists. One of the most consistent of the legal critics is the distinguished professor of jurisprudence, Edmond Cahn, of New York University Law School. Ernest van den Haag is an example of a critic from within the field of social sciences. The bulk of this paper will be devoted to an analysis of the criticisms of Professor Cahn and Dr. van den Haag because Professor Cahn has undoubtedly influenced the thinking of other students of jurisprudence[1] and Dr. van den Haag has presented the most specific and intense critical comments that have so far been published by a social scientist.

The criticisms of Professor Cahn take many forms.[2] Essentially, however, he states that it is incorrect to believe that the Brown decision[3] was "caused by the testimony and opinions of the scientists" and that the constitutional rights of Negroes or any other Americans should not "rest on any such flimsy foundation as some of the scientific demonstrations in these records." He contends that the cruelty inherent in racial segregation "is obvious and evident."

Among his other charges are: (1) that this writer exaggerated the contribution of social science experts to these cases; and (2) that in writing a report of the role of social scientists which was published before May 17, 1954, the writer could not have known that Chief Justice Warren's opinion would not mention either the testimony of the expert witnesses or the statements submitted by the thirty-two social scientists. Professor Cahn added solicitously:

> "The Chief Justice cushioned the blow to some extent by citing certain professional publications of the psychological experts in a footnote, alluding to them graciously as 'modern authority.' In view of their devoted efforts to defeat segregation, this was the kind of gesture a magnanimous judge would feel impelled to make, and we are bound to take satisfaction in the accolade."

In speculating on why the Court did not mention the social scientists' brief in its opinion, Professor Cahn states his personal, subjective reaction that the text of his statement conveyed little or no information beyond what is known as "literary psychology." The fact is, however, that all but one of the references cited by the Court in footnote 11 of the Brown decision were cited as references in the social science brief which had been submitted to the Court. The one reference which had not been listed but cited by the Court was Witmer and Kotinsky's Personality in the Making, the relevant portion of which was a summary of this writer's White House Conference manuscript on the effects of prejudice and discrimination on personality development.

Whatever might be one's degree of agreement or disagreement with Professor Cahn's estimate of the worth of the social scientists' testimony in these cases or the degree of the Court's regard for the social scientists' material presented in the brief or in the trial records, one must take seriously his argument that the constitutional rights of Negroes or other Americans should not rest on social scientists' testimony alone. If he had concentrated and elaborated on this issue on a high level of academic discourse, he might have made an important contribution to thought in a field in which he is competent. When he leaves the area of the law, constitutional rights and matters of jurisprudence and invades the area of social sciences, making broad and general comments about the validity of social science methods, premises, approaches, findings and conclusions, and when he explicitly or implicitly attacks or suggests that the social scientists who participated in these cases as witnesses and consultants did not do so with the utmost personal scientific integrity, he gratuitously leaves his field of competence and communicates his personal opinions, biases and misconceptions as if they were facts. His prestige in a field in which he has been trained thereby disguises his ignorance in a field in which he has no training. For these reasons, it is necessary to answer these charges and generalizations with clarity.

Some Relevant Facts.

Before one enters a general appraisal of the validity of some of the many assumptions, implications, and charges raised by Professor Cahn, it is necessary to clarify certain points of fact which are relevant to opinions about the role of social scientists in these cases:

(1) The social scientists who participated in these cases were invited to do so by the lawyers of the NAACP. It was these lawyers who had the primary and exclusive responsibility for developing the legal rationale and approach upon which these cases would be tried and appealed. It was they who made the decision to bring the legal attack on the problem of overruling the Plessy "separate but equal" doctrine[4] by attempting to demonstrate that state laws which required or permitted segregation in public schools violated the equal protection clause of the fourteenth amendment. It was their decision that the chances of success would be greater if it could be demonstrated that racial segregation, without regard to equality of facilities, damaged Negro children. Furthermore, it was their decision to determine whether they could find acceptable evidence from social psychology and other social sciences which would support their belief that psychological damage resulted from racial segregation. Social scientists were not involved and did not participate in any way in these initial and important policy or legal strategy decisions. Only after these decisions were made by the lawyers of the NAACP were the social scientists approached and invited by the lawyers to participate in these cases. The social scientists were asked whether there were any relevant scientific studies on the psychological effects of racial segregation. Finally, it was the judgment of these lawyers that the studies and evidence offered by the social scientists were relevant and crucial enough to form an integral part of their trial and appellate case.

(2) The studies which were relied upon by the social scientists in arriving at the conclusion that racial segregation damaged the human personality were not studies which were conducted specifically for these legal cases. Systematic research on the psychological aspects of racial prejudice, discrimination, and segregation had been going on for more than fifteen years. The White House Conference manuscript, which was cited by the United States Supreme Court in footnote 11 in the Brown decision, was a compilation of all of the available knowledge of the effects of prejudice and discrimination on personality development in children and was prepared by this writer months before he was aware of the fact that the NAACP intended to bring cases before the federal courts challenging the validity of segregated schools.

(3) The studies cited in this White House Conference manuscript and the joint primary research of this writer and his wife formed the bulk of his testimony in three of these five cases. The primary research studies were conducted ten years before the cases were heard on the trial court level. Professor Cahn's allegation that the writer served in the role of advocate rather than that of an objective scientist in this participation in these cases seems

difficult to sustain in the face of testimony given on the basis of research conducted ten years before these cases were heard. One would have to be gifted with the power of a seer in order to prepare himself for the role of advocate in these specific cases ten years in advance.

(4) The use of the "Dolls Test" (actual dolls, not pictures of dolls, were used in this research) on some of the plaintiffs was to determine whether the general findings from the larger number of Negro children who had been tested years before were true also for the children who were the actual plaintiffs in these cases. The decision to test some of these plaintiffs was a legal one made by the lawyers of the NAACP. It was their assumption as lawyers that general scientific findings would have more weight in a courtroom if it could be demonstrated that they also applied in the specific cases and for the particular plaintiffs before the court. When these plaintiff children were tested and interviewed by this writer, it was his judgment that some of these children showed evidence of the same type of personality damage related to racial prejudice, segregation, and discrimination which was found in the larger number of subjects who were studied in the original, published research. This opinion was presented to the courts in the form of sworn testimony.

(5) The justices of the federal district courts were at all times free to rule that the testimony of the social scientists was irrelevant and immaterial. The United States Supreme Court could have refused to accept the Social Science statement which was submitted to it in the form of an appendix to the legal brief of the appellants. If either of these had been done, there would now be no question of whether the courts did or did not reply on the findings and opinions of social scientists.

It is still a matter of social reality that social scientific findings and opinions are not incorporated into, nor do they determine, policy decisions, legislative action, or judicial decisions except to the extent that those who have the power to make these practical decisions choose to accept or reject the relevant findings of scientists. Whether this should continue to be so is, of course, debatable.

"Fidelity," "Truth," and Academic Courtesy

Professor Cahn implies that the primary motive of the social psychologists who participated in these cases was not "strict fidelity to objective truth." This is a serious, grave, and shocking charge.

Professor Cahn did not present evidence to support his implication that the social scientists who participated in these cases, and particularly this writer, betrayed their trusts as scientists. He merely makes the assertion that some day judges will be wise and will be able to notice "where objective science ends and advocacy begins." For the present, however, "it is still possible for the social psychologists to 'hoodwink' a judge who is not overwise...."

It is difficult to take this type of comment seriously. Since it has been published over the signature of an individual who commands the respect of his legal colleagues, it cannot be dismissed. It cannot be waived aside as evidence that Professor Cahn believes himself wiser than the entire legal staff of the NAACP, the battery of lawyers employed by the opposition--including the late John W. Davis, who devoted a considerable amount of space in his Supreme Court brief and in his first arguments before the United States Supreme Court to the social science testimony--or the lawyers of the Department of Justice of the United States, and, finally, the Justices of the United States Supreme Court.

This point must be answered by a description of concrete facts in the relationship between the NAACP lawyers and the social scientists who were involved in these cases. The social scientists who testified in these cases or endorsed the Social Science appendix at the invitation of this writer were not the type of human beings who were capable, personally or professionally, of testifying to a fact or stating an opinion which they did not believe to be consistent with the scientific evidence as they knew it. These men are neither infallible nor all-wise; but they are the outstanding experts in this field. What is even more important, they are men of integrity.

When the lawyers of the NAACP, in their understandable zeal to develop the strongest possible case, asked the social scientists whether it was possible to present evidence showing that public school segregation, in itself, damaged the personalities of Negro children, it was pointed out to them that the available studies had so far not isolated this single variable from the total social complexity of racial prejudice, discrimination, and segregation. It was therefore not possible to testify on the psychologically damaging effects of segregated schools alone. Such specific evidence, if available at all, would have to come from educators and educational philosophers. Some of the more insistent lawyers felt that only this type of specific testimony would be of value to them in these cases. It was pointed out to these lawyers that if this were so then the social psychologists and other social scientists could not be of any significant, direct help to them. A careful examination of the testimony of the social scientists, found in the record of these cases and the Social Science appendix submitted to the United States Supreme Court, will show that the social scientists presented testimony, opinions, and information consistent with the available empirical studies, conclusions, and observations. They presented this information with caution and restraint befitting their roles as trained and disciplined scientists. As expert witnesses, they made not a single concession to expediency, to the practical and legal demands of these cases, or even to the moral and humane issues involved as they adhered to their concept of "strict fidelity to objective truth." Certainly Professor Cahn cannot be the judge of whether his concept of "strict fidelity to objective truth" in the field of social science is more acceptable or valid than theirs.

It must also now be stated that one of the responsibilities assigned to this writer in his role of social science consultant to the legal staff of the NAACP was to advise the lawyers not only about those studies and individuals who were scientifically acceptable, but also to advise and warn them away from studies and individuals of questionable scientific repute. At least one well-publicized report on the damaging effects of segregation on the personality of Negroes was not used in these cases because it was the judgment of this writer, which was communicated to and accepted by the lawyers, that its methodology was scientifically questionable, its selection of subjects and sampling were clearly biased, and that its conclusions bordered on the sensational. In short, it was believed that in spite of the fact that this study purported to present clear evidence in support of the hypothesis that racial oppression damaged the personality of Negroes, its flaws and scientific inadequacy were so clear it could not be defended in court.

It is difficult to determine precisely what Professor Cahn means by "objective truth." According to his article "most of mankind already acknowledged..." that segregation is cruel to Negro children, involves stigma and loss of status, and may ultimately shatter their "spines" and deprive them of self-respect. The "shattering of spines" is Professor Cahn's contribution to the knowledge of the detrimental effects of racial segregation. No social scientist testified to this "fact." Professor Cahn contends, however, that when scientists attempt to demonstrate these same "well-known facts" through their use of the methods and approaches of science, they "provide a rather bizarre spectacle." What is more, he maintains they exaggerate their role, their methods are questionable, their logic and interpretation weak and fallacious, and they distort their findings as they become advocates who seek to "hoodwink" the judges. A serious question would be: How could the social scientists be so unreliable yet nonetheless come out with a picture of social reality which Professor Cahn and everyone else "already knew"?[5]

Professor Cahn presents a novel concept of the relationship between common knowledge and scientific knowledge. The logic of his position rests upon the premise that science concerns itself with one order of reality which is distinct from other forms of reality or truth--that a scientific "fact" has different attributes or characteristics than a "fact" or common knowledge. Another related theme which runs through his comments is that a "legal fact" is distinct from both a "scientific fact" and a "fact of common knowledge."

Cahn's pluralistic approach to the nature of "facts," while not a novel philosophical position, seems to involve a mystical semantic confusion which is inconsistent with the assumptions imperative for a scientific approach to the understanding of the nature of man, his society, and his environment.

Science is merely the last of many approaches that man has used in his attempt to determine the "facts" and truth of nature. As the late Professor Einstein has observed: "Scientific thought is a development of pre-scientific thought." Before and coincident with science, man tried mysticism, religion, and philosophy in his attempts to determine the facts of nature. In his quest to control his environment and his relations with his fellow human beings, he attempted to implement his various types of "knowledge" by seemingly compatible techniques of control, e.g., magic, prayers, reason, law, and technology. These various approaches in the quest of truth and the control of the environment were not seeking different types of truth. Indeed, it must be assumed that science and technology developed precisely because earlier approaches to the nature of "truth" and "fact" left much to be desired by way of successful demonstration of the practical utility or the human consequences of these "truths" and "facts."

The development of science as an approach to the determination of truth involved the development of methods for the control of errors in human observation, judgment, biases, and vested interests. These were the factors which seemed to have distorted man's concept of, or blocked his contact with, the "truth" or "facts" of experience. When they are operative, man's "common knowledge" becomes inconsistent with "scientific knowledge." When they are controlled or for some other reason non-operative, "common knowledge" and "scientific knowledge" are coincident--both reflecting the nature of reality, truth, or facts, as these are knowable to the human senses and intelligence.

Science is essentially a method of controlled observation and verification for the purpose of reducing human errors of observation, judgment, or logic. Science begins with observation and ends by testing its assumptions against experience. It is not a creation of another order of reality. In a very basic sense there cannot be a "legal fact" or a "fact of common knowledge" which is not at the same time a "scientific fact." Whenever this appears to be true, one or the other type of "fact" is not a fact.

Kenneth B. Clark

The Basic Issue

After one has cut through the emotional irrelevancies of Professor Cahn's article, one is confronted with the basic circuitous plea that the law and the courts of the land should be isolated in Olympian grandeur from the other intellectual and scientific activities of man. Specifically, Cahn seems primarily--even if unconsciously--disturbed by the fact that the upstarts of the new social sciences should have been involved at all in these important cases which belonged exclusively to lawyers and students of jurisprudence. It is to be hoped that a decreasing number of lawyers believe that laws and courts are sacred and should be kept antiseptically isolated from the main stream of human progress. Such isolation cannot be and never has been true except in the classrooms of some puristic law school professors.

The law is concerned with society and the regulation of human affairs. Social science, government, philosophy, and religion are also concerned with society, its understanding and regulation. Man's relations with his fellow man involve matters far too grave and crucial to be left to lawyers and judges alone. Respect for the law, intelligently and ethically conceived and executed, is essential for stable government. Intelligence and ethics cannot stem from the law alone but must be fed to it through the ceaseless struggles of scholars, scientists, and others toward truth and understanding. This may be difficult for Professor Cahn to accept. It nonetheless remains a fact.

As Brandeis once said: "A judge is presumed to know the elements of law, but there is no presumption that he knows the facts."[6] With the vast range and types of cases which come before the courts, it is unlikely that even the wisest judges and lawyers could be competent in all fields of human knowledge. One may presume that it was a recognition of these facts among others that influenced the decision of the lawyers of the NAACP to seek the help of social scientists in their attempt to overrule the Plessy v. Ferguson "separate but equal" doctrine which had dominated civil rights litigation since 1896.

Another important fact which was ignored by Professor Cahn in his castigation of the social scientists' role in these segregated school cases was the fact that this was not the first time that the lawyers of the NAACP had sought to convince the United States Supreme Court that segregation in and of itself was unconstitutional. In the Sweatt[7] and McLaurin[8] cases they sought a decision on the issue of segregation per se by relying on the traditional legal approach. Substantially the same United States Supreme Court which handed down the

Brown and Bolling[9] decisions, however, decided the Sweatt and McLaurin cases within the framework of the Plessy "separate but equal" doctrine. It may merely be coincidental that the lawyers of the NAACP succeeded in overruling the Plessy doctrine only after they enlisted an impressive array of social science testimony and talent and attacked this problem with this approach.

Another Point of View.

Some astute students of jurisprudence hold opinions on this issue which differ from those presented by Professor Cahn. The late Alexander Pekelis, in making his case for a jurisprudence of welfare, stated:

> "A great many contemporary judicial decisions show this threefold leitmotif--awareness of freedom, confession of fallibility, and quest for extra-legal guidance....
>
> "A participation of the social sciences in the development of a welfare jurisprudence may bring the normative elements in social science into the light of consciousness, and thus contribute to a healthy development of social theory....
>
> "The economic and social facts of life, which legal realism has taught us, have banished the belief that judicial decisions are brought ready-made by constitutional storks.... Similarly, society cannot be built upon judicial whim or expediency alone.
>
> "We cannot turn back the clock. Social scientists (economists, sociologists and psychologists) are with us for good, and are going to remain in the very midst of government.... Judges may and should become acquainted with the various non-legal disciplines.... A judge should know more about social studies precisely in order to acquire the conviction that they can furnish no more certainty than constitutions, statutes or precedents."[10]

It would be fatuous to argue that because there is difference of opinion among eminent students of jurisprudence that, therefore, judicial opinions should not be taken seriously. Of course there are dangers involved in the use of science in any area of human activity. There are undoubtedly some social scientists who might be willing to sell their intelligence, training, and themselves to the highest bidder. There are those who will be easily intimidated by the practical demands of vested interests and men of power. There are those who will rationalize their subservience by demonstrating their affluence and tough-minded practicality --or even their scientific purity. But this is not new. Science has nonetheless continued its advance and contributions to the ethical and material progress of mankind.

FOOTNOTES

*Associate Professor of Psychology, City College of New York; co-founder and Director of Research for Northside Center for Child Development, New York City; Social Science Consultant to the legal staff of the NAACP. A.B. 1935, M.S. 1936, Howard University; Ph.D. 1940, Columbia University.

[1] See BLAUSTEIN AND FERGUSON, DESEGREGATION AND THE LAW 135-37 (1957).

[2] Cahn, Chapter on Jurisprudence, *Annual Survey of American Law*, 30 N.Y.U. L. Rev. 150-69 (1955).

[3] Brown v. Board of Educ., 347 U.S. 483 (1954).

[4] Plessy v. Ferguson, 163 U.S. 537 (1896).

[5] It may be noted parenthetically that it is questionable whether all judges share this "common knowledge," as is evidenced by the prior decision that upheld the "separate but equal" doctrine. At any rate, Professor Cahn does not explain why these judges did not act upon their knowledge. In fact, he does not explain why these judges did not act upon their knowledge. In fact, he does not explain how a person not gifted with superior insights can determine what is and what is not "common knowledge" as distinct from the personal biases of judges. Nor does Professor Cahn suggest any means, other than through the medium of expert witnesses, for getting "common knowledge," critically examined, into the court record so that it may be considered by judges who have the responsibility for the final decision.

[6] MASON, BRANDEIS, A FREE MAN'S LIFE (1946).

[7] Sweatt v. Painter, 339 U.S. 629 (1950).

[8] McLaurin v. Oklahoma State Regents, 339 U.S. 637 (1950).

FOOTNOTES

[9] Bolling v. Sharpe, 347 U.S. 497 (1954) (decided the same day as the *Brown* decision). The Court held segregated schools in the District of Columbia to be unconstitutional.

[10] Pekelis, *The Case for a Jurisprudence of Welfare*, 2 SOCIAL RESEARCH No. 3, reprinted in 6 LAWYER'S GUILD REVIEW No. 5 (1946).

LAW AND SOCIAL SCIENCE METHODOLOGY: THE NEED FOR ASSIMILATION

Not only does Hans Zeisel's article present a panorama of social science studies being used by the legal profession; it also implicitly suggests that if the law is to make good use of such studies, the methods underlying such studies must be assimilated into the law's disciplinary framework. Otherwise, lawyers are all too likely to be forced into blindly accepting whatever is handed them.

Hans Zeisel, "The Law," in The Uses of Sociology, eds. by Paul F. Lazarsfeld, William H. Sewell, and Harold L. Wilensky (New York; Basic Books, Inc., 1967), pp. 81-99.

The law from one view is a continuous process of synthesizing facts and rules, with new facts at times engendering new rules. The process takes place wherever law is made: in the legislatures, in the administrative agencies, and in the courts.

The facts reach these lawmakers in a variety of forms. The primary source is still the witness who reports on his own private experience. Occasionally, however, facts are presented as cumulative knowledge, systematically gathered through surveys and most recently also through experiments, methods that are part of the tool chest of the social sciences.

The survey as a source of facts for the law predates modern social science by centuries and constitutes in fact one of its major historical roots.[1] But it is only in recent years that the law has begun to use research operations conducted with technical rigor. The uses the law has made of such systematic investigations differ widely, from simple citation in a brief or opinion to being the decisive ground for a judgment or a legal reform.[2] The great majority of the studies that come before the law raise only private issues, assisting courts and agencies in individual litigation. They may be surveys of the quality of contracted goods, of the geographic range from which a drive-in theater draws its clientele, of the commercial effects of a merger, or of the socioeconomic structure of the jurors in a certain community. But although such studies are at times gems of technical perfection and ingenuity, they will not be discussed here. Their variety is too great, and they seldom reach the higher courts, hence they seldom affect legal rule making; and they have been sufficiently discussed elsewhere.[3]

The studies we shall discuss here deal with more general problems: with substantive rules of law, with procedural rules, or with institutions that are a mixture of both. Some of these studies are broad surveys of a legal institution without more specific focus; they enter the stream of legal resolution only slowly, as but one of the many sources that shape the law. Other studies, in contrast, are designed to illuminate if not to resolve one narrow, crucial issue; these investigations—sometimes they are controlled experiments—are bound to affect the law more directly.

But all the studies on which this chapter will report have one thing in common: they were made for the purpose of being used by the law and in many instances have affected its course.[4]

We shall report these studies in the order of the spectrum that ranges from the narrowly but sharply focused controlled experiment to the broad, diffuse survey.

Controlled Experiment

There is no more powerful tool for assessing a legal innovation, or any innovaton for that matter, than the controlled experiment. But since its essence is to apply a rule of law to one group of cases and to withhold it from another, such purposeful discrimination would seem at first to violate the equal-protection guaranty of the Constitution.[5] Indeed, there are substantive limits to legal experimentation; there can be none that involves withholding of a right, guaranteed under all circumstances. It would be impossible, for instance, to insist that some criminal defendants be tried without counsel in order to find out what effect counsel has. But there are several reasons why, outside this rigidly protected sphere, the law should permit experimentation. First, the experimental discrimination is by definition temporary; second, the discrimination is applied impartially, by lot; third it is the very purpose of the experiment to learn what, if any, effect the discriminating rule would have: hence, at the outset, one cannot even be certain that there is discrimination; and fourth, the ultimate aim of the experiment is to eliminate the rule if it should be found to discriminate unfairly.

Controlled legal experiments, not surprisingly, have largely been confined to rules that convey privileges rather than rights.

The first controlled experiments within the precincts of the law were probably conducted by the Adult Prison Authority of the State of California, which tried to assess the effectiveness of a variety of prisoner treatments. The most daring of these experiments was an effort to determine what happened if prisoners were released nine months before their appointed time; specifically, whether such a premature release was likely to increase the rate of recidivism.[6]

Not much came of these experiments, partly perhaps because the recidivism rate is too brittle a measure of effectiveness since, in order to be counted as a recidivist, it is not sufficient to have committed another crime; it is also necessary to be caught and reconvicted. And since the odds of being caught, as revealed by the published statistics, are on the average about one in five, this ratio might well have a great variance and hence be a very unstable measure.[7]

A controlled experiment that allowed of precise measurement and had an immediate effect on the law was conducted in the state courts of New Jersey.

In most of our courts some or all of the civil suits, before they come to trial, are scheduled for what has become known as pretrial. There, counsel for both sides, occasionally with their clients, meet with the judge to present briefly the issues under dispute and air the possibilities of settlement. Tradition has it that these pretrials, aside from preparing and facilitating the subsequent trial, increase the rate of settlements prior to trial. The institution has, therefore, been considered a most desirable means of reducing the trial load and thereby the intolerable congestion of our metropolitan courts. Since many cases, the trial of which would have lasted two days on the average, are settled during a half-hour pretrial conference, this notion seemed well supported.

But analysis of available statistics made the point doubtful; there were indications that the cases settled at pretrial would have been settled even without it and that the court time spent on pretrying cases might be wasted. The precise answer, it was suggested, could come only from a controlled experiment which pretried a random sample of cases and omitted pretrial in a comparable control group.[8]

At that time the New Jersey courts had a rule that made pretrial obligatory, and the state's distinguished Chief Justice and its Court Administrator, becoming sympathetic to both the query and the proposal, commissioned Professor Rosenberg, then Director of the Project for Effective Justice at Columbia University, to conduct the experiment.[9] The design called for random assignment of cases by the clerks of the respective courts to two alternative procedures: to obligatory pretrial in one group of cases, and to optional pretrial in the control group, where it would be held only if one or both of the litigants requested it.[10] Accordingly, 2,954 cases were assigned at random alternatingly to the two groups, for which the settlement ratios shown in Table 4–1 emerged. There was no difference. In addition, the experiment failed to confirm a subsidiary hypothesis, namely, that the pretried, and hence prepared, cases required a shorter trial time. Thus the conclusion emerged simple and clean: contrary to a widely held belief, obligatory pretrial did not save court time, but in fact wasted it.

TABLE 4-1.

	CONTROL GROUP: OBLIGATORY PRETRIAL	EXPERIMENTAL GROUP: OPTIONAL PRETRIAL
Suits settled before they reached the trial stage	76%	78%

Persuaded by the experiment, the State of New Jersey forthwith changed its rule and made pretrial optional.[11]

But the legal experiment that had the most profound and sweeping effect on the law was conducted in the criminal courts of Manhattan. It revolutionized one of the most solid traditions in the criminal law: the practice of setting bail for defendants arraigned in our criminal courts. Bail is set, as a matter of constitutional right, for nearly all defendants; if they can post it, they are set free; if not, they must remain in jail. Whether or not they *can* post it depends as a rule on the bondsman, who, against a premium of some 10 per cent, will or will not take the risk of providing the demanded bail. Only rarely is a defendant allowed to go free without posting bail.

The system has been heavily criticized because it favors the well-to-do, surrenders the actual decision to the bondsman, and keeps an inordinate proportion of defendants in jail, some of whom are subsequently acquitted. The system, nevertheless, withstood all criticism until the Vera Foundation made known its findings from a unique experiment which it conducted in 1961. With the co-operation of the New York judiciary and the New York University Law School, all defendants arraigned in the felony court of Manhattan were interviewed so as to assess the risk of their failing to appear at their trial, if the court were to free them without requiring bail. On the basis of these interviews the defendants were classified into two groups: those for whom a release without bail could be reasonably recommended to the court and those for whom such a recommendation could not be made. The recommendable group was then divided into two random halves: the experimental group, for which the recommendation to release the defendant without bail was actually transmitted to the arraignment judge, and the control group, with respect to which the judge was told nothing and thereby left to his own traditional mode of making the bail decision. In this latter group only 14 per cent of all defendants were freed without bail, as against 60 per cent in the recommended half. The hypothesis was that at the time of trial, from the group of which 60 per cent had been freed without bail, more would fail to appear in court than from the group where only 14 per cent were free without bail, and the question was: how many more? When trial time came, only 1 per cent of *all* defendants released without bail, whether recommended or not, purposely failed to appear in court. The experiment thereby proved that the number of defendants released without bail could be quadrupled without reducing their availability at the time of trial.[12]

The results of the experiment were stunning. The City of New York took over the interviewing from the foundation and established it as a permanent service. The Attorney-General of the United States convoked a conference on the topic, and today almost all major cities and many rural areas have adopted the Vera procedure, and with it the liberalized practice of release without bail. And the Department of Justice left no doubt as to where the credit belonged: "Of particular significance is the fact that these changes have flowed not out of a crisis . . . but rather from education, through empirical research and demonstration." [13]

The secret of the success of this experiment was twofold. First, except for the bail bondsmen, everybody stood to gain from the liberalization: the municipal jails saved money; the defendants themselves were spared unnecessary hardships; and last, but not least, the ends of justice were advanced. Secondly, the numerical result of the experiment was so clear that no probability calculus was needed for its appreciation.[14]

The Natural Experiment

Sometimes administrative routine will present the investigator with a natural experiment, that is, with an experimental and a control group that were not purposely designed by him.

The study of differential sentencing provides a classic example. The law has very little control over the sentence of a convicted defendant. It merely sets the range, usually wide, and thus leaves much discretion to the judge.[15] Since many courts assign cases at random to the different judges, the cases before each judge form a natural experimental unit. Thus, a study of the New York City Magistrate Court revealed that, among the thirteen judges, one discharged 73 per cent of the defendants before him on a charge of intoxication, while one of his colleagues discharged only 1 per cent.[16] Later studies, especially those of Gaudet, confirmed the problem.[17] Today the disparity of sentences for comparable crimes has become a major concern of the judiciary. Federal judges have instituted procedures designed to minimize these fluctuations.

A slightly different, natural experimental situation is provided by collegiate judicial tribunals, inasmuch as the judges there are confronted by the identical court record in each case, yet often come to different decisions. These studies often employ rather sophisticated analytical techniques, such as scale and factor analysis. But in some cases the data are so stunningly lopsided that they can stand in the raw. Thus it was found that in the Michigan Supreme Court's decisions on claims under the Workmen's Compensation law the judges simply split along party lines.[18]

Pritchett was the first to analyze collegiate court decisions in his studies of the United States Supreme Court.[19] In recent years studies of this type have proliferated.[20] The reason why they are not given more prominence in this essay is that they have yet proved of little "use" in the meaning of this book.

Following we report on a study of a different type. It too made use of variations among judges. But, in contrast to the studies cited above, it was a study which the court itself initiated to provide guidance for its procedural rules.

Under traditional trial procedure, the plaintiff in a civil case first presents his case both as to liability and size of damages and is followed in turn by the defendant, who presents his side of the case. After both have had their say, the jury retires and decides whether the defendant is at all liable for damages and, if so, how large these damages should be. The question of damages thus becomes relevant only if liability is found. The suggestion was made to split the trial and to limit evidence and argument in the first part of the trial to the liability issue, asking the jury to decide whether the defendant owes anything at all. Only if this decision is affirmative does the trial proceed with the evidence and subsequent verdict on the *size* of the damages.[21] Since liability is affirmed in only a little more than half of all cases, this mode of trial was expected to save something like half of the trial time normally spent on damages. The Federal District Court for Northern Illinois was sufficiently intrigued by this split-trial idea to try it out and to ask the University of Chicago Law School to help in assessing the effect of the split-trial rule, as it has come to be called.[22]

The court had adopted the rule in a form that left it to the discretion of the individual judges whether they wanted to apply it in the particular case. It was from this discretion that, seemingly, the difficulties, but eventually the salvation of the experimental design, arose. If each judge could apply the rule in some cases and not in others, and if he were to select—as in fact he often did—only those cases for split trial which, in his view, promised some gain in time from the application, the cases tried in the regular mode and those tried under the split-trial rule could not be compared. Whatever difference in trial length might be found between the two could not be attributed to the new rule, because the cases were admittedly different to begin with.

At first glance this lack of random assignment would seem fatal. Yet, while it made the analysis more complicated and less powerful, it did in fact make the experiment possible. Since the original assignment of cases to the individual judges was made randomly, the inference was allowed that the cases coming before Judge A did not differ from the cases coming before Judge B. And then something fortunate happened. The discretion of the judges resulted in an effective spread of the experimental stimulus: some used the rule in almost all their cases, some in hardly any, and some in varying proportions between.

If, then, it were true that the application of the split-trial rule saved time, the judges who applied the rule more often should require on the average less trial time than those who applied it less often. This turned out to be true, as Table 4–2 shows.

TABLE 4–2. *Proportion of Split Trials and Average Trial Time in Personal Injury Trials*

JUDGE	PROPORTION OF CASES TRIED UNDER SPLIT-RULE (PER CENT)	AVERAGE LENGTH OF ALL TRIALS BEFORE THIS JUDGE (DAYS)	NUMBER OF TRIALS BEFORE THIS JUDGE*
A	80	3.2	(26)
B	51	3.3	(41)
C	38	3.5	(26)
D	14	3.8	(22)
E	7	3.9	(27)
F	7	4.3	(14)

* Only judges with more than 10 trials are included.

The regression line based on these data indicated that at the point where a judge conducted all trials under the split-trial rule, his average trial time was about 20 per cent below the point where none of the trials were split. This, then, was the magnitude of the time that could be saved through application of the rule.

But since the variation in the stimulus was not random, but self-selected, it was desirable to provide supporting evidence. It was clear that whatever savings there were must come from the elimination of the damage trial. The frequency of damage trials was, therefore, determined both for the regular and for the split trials. (See Table 4–3.)

The dispositions of the two groups of cases were drastically different: 76 per cent of all regular cases went through a complete trial, against only 15 per cent of the separate trials. As for the latter group, 58 per cent were spared trial of the damage issue because of the intermediate verdict denying liability. In 43 per cent of the cases the trial simply ended when liability was denied, and in another 13 per cent of the cases there was no trial of the damages because they were settled after the jury had affirmed liability.

The split trial is a radical innovation in American law, brought to the fore by the pressures of court congestion. When the study was first pub-

TABLE 4–3. *Disposition of Case in Regular and Separated Trials*

	REGULAR TRIALS (PER CENT)	SEPARATE TRIALS (PER CENT)
Complete trial on liability and damages	76	15
Trial ended after liability verdict	—	58
because verdict was for defendant	—	43
because damages were settled after verdict affirming liability	—	15
Other dispositions (settlement during trial, directed verdicts)	24	27
	100%	100%

lished, the Joint Committee on Effective Administration of Justice, on suggestion of its chairman, United States Supreme Court Justice Clark, had copies sent to all trial judges in the United States; as of this writing, several state and federal courts have moved to authorize their judges to use the split-trial procedure, and the institution is likely to gain more ground.[23]

Experiment under Seminatural Conditions

If an issue falls into the constitutionally protected area, not even a natural experiment is likely to occur, and a simulated one must suffice. But even then, as many natural components as possible must be retained, as for instance in the following series of experiments concerning the defense of insanity. They were designed to test jury reaction to certain variations in the law. The natural element in these experiments was the jurors, summoned by a real trial judge from the jury pool of his court with the request to partake in the experiment and to deliberate on the case as if it were a real one.[24]

In Anglo-American law the defense of insanity has been embodied for more than a century in the so-called M'Naghten rule, which calls for an acquittal if the defendant either did not know what he was doing or did not know that what he was doing was wrong. Recently, the rule has come under criticism, primarily from psychiatrists. In 1954 the Federal Court in the District of Columbia established a new rule in a case in which one Durham was indicted, and subsequently acquitted, on a charge of burglary. The Durham rule considers the defense as established if the criminal act can be shown to be the "product of a mental disease or a mental defect." It became thus a point of major interest for the criminal law to find out what if any difference it made to the outcome of a trial whether insanity was defined under the M'Naghten or under the Durham rule.

The "law" in a criminal jury trial becomes operative primarily through the judge's instruction to the jury before it begins deliberation. In that instruction the judge spells out the circumstances under which the jury may find the defendant insane. In a way, then, the question as to what difference the law makes means what difference it makes to the jury whether it is instructed according to the rule in M'Naghten or in Durham.

To compare the insanity cases in the District of Columbia with cases from a court that operates under the M'Naghten doctrine could provide only unsatisfactory findings, since not only the rule of law but also the cases, the juries, and the judges are likely to be different. And it is obviously impossible to decide the issue through a controlled experiment under completely natural conditions. Therefore, an experiment had to be designed that combined natural with laboratory conditions, sufficiently realistic to justify confidence in its validity.

Two trial records were composed: one, a case of housebreaking, a simplified version of the original Durham trial; the other, an incest case, also an abbreviated version of an actual trial. In both trials the accused's only defense was insanity. The trial evidence was acted out and with the other elements of the trial put on recording tape. Of each case, three main variants were produced.[25] The tapes were identical but for that part of the judge's instruction that dealt with the defense of insanity and for the concomitant psychiatric testimony. In one version the instruction and psychiatric testimony were according to M'Naghten; in the second, according to Durham; and in the third, the instruction left it in fact to the jurors own judgment as to whether the evidence in the case supported a defense of insanity, forcing the jury to establish its own law of insanity.

Each of the three versions was then taken into two metropolitan courts and presented in turn to more than a hundred juries. A judge called these jurors into his courtroom and asked them to co-operate in the experiment; by so doing, he advised them, they would oblige the court and also discharge their present turn of jury duty. The jurors then listened to the taped trial and afterward deliberated and arrived at a verdict. Table 4-4 shows the outcome of the experiment in terms of the jurors' vote on their first ballot, prior to the beginning of the deliberation.[26]

In both trials, the Durham rule elicited a higher percentage of acquittals by reason of insanity than the M'Naghten rule. That the percentages

TABLE 4-4. *Per Cent of Jurors Voting "Not Guilty—Insane" on First Ballot*

	M'NAGHTEN	DURHAM	NO RULE
Incest case	24 (240)	36 (312)	34 (264)
Housebreaking case	57 (120)	65 (120)	76 (120)

under Durham are very close to those obtained under the "No Rule" instruction suggests, furthermore, as indeed it has been argued,[27] that Durham comes close to being no rule.

The figures show that the incest case allowed a sharper differentiation between M'Naghten and Durham than the housebreaking case. This difference is instructive beyond the specific issue. The defendant in the incest case was an officer in a city's fire department, with an excellent record, who, except for the crime in question, had never shown any signs of abnormality. The defendant in the burglary case, on the other hand, much like the original Durham, had been in and out of mental institutions and hence had shown, by whatever legal or common-sense rule, signs of insanity. The fireman could be found insane only if the jury was instructed (as Durham allows) to consider the criminal act itself as a symptom of insanity.

The experiment raised two questions of general significance. One applies to the realism of simulation; the other, to the degree of generalizing from experimental findings.

In this experiment there were two simulated elements: the experimental stimulus was greatly reduced (a two-day trial condensed into an hour), and jury deliberation was clearly a mock procedure without consequences in the real world. As to the second point, there was considerable reassurance: these were real jurors called to duty by a real judge; they discharged their responsibilities with such obvious zeal and honesty that deliberations lasted up to ten hours, often engendering high-pitched battles among the jurors who, at times, ended in a "hung jury."

The other point raises a more serious problem. A tape recording of a trial, condensed to about an hour's length, may be something quite different from a full-blown trial, in which the jury not only hears but also sees over a period of many hours real people with all the significant details of their reactions. Without additional research, the point allows of no precise answer.

As to the consequences of the insanity experiments, one can at this stage only venture a guess. On the whole, the experiment should strengthen the hand of those who oppose the Durham rule, simply because its message to the jury is ambiguous. Whatever the shortcomings of the traditional right-wrong test, its criteria are clear and can be applied by the jury.

The Survey Experiment

The jury, in spite of its deep constitutional roots, has been the topic of perennial debate, with little precise knowledge to support it.

Some years ago, the University of Chicago Law School began a large-scale study of the jury system, and one of its key questions was: What difference would it make if all jury cases were tried only by a judge sitting without a jury?

The question would seem to demand a controlled experiment—every case to be tried twice, once with and once without a jury—an obviously impossible solution. Equally impossible it would be to assign cases at random to jury and judge, since this is a choice no defendant must be deprived of. Nor would the simple comparison of actual jury verdicts with actual judge verdicts help, even if limited to trials of the same type of crime, because we know that the cases in which the defendant waives a jury are quite different from those where he wants one.

Curiously enough, the design eventually adopted for the study came close to the ideal design, the controlled experiment. A nationwide sample of trial judges reported for a specified time period on all the jury trials over which they presided. Each judge told us how the jury decided the case and how he, the judge, would have decided it, had he sat without a jury. The design thus made use of the fact that every case is tried twice, albeit simultaneously: once before the jury and once before the presiding judge, who, if there were no jury in the case, would have to render the judgment.

This research design is but a natural controlled experiment of a special order. Experimental stimulus (the jury) and control (the judge) are present in every case, but unlike a planned experiment, there is, except for the judge who may preside over several trials, no replication. The jury changes from trial to trial; and most important, the case, too, of course, is never the same. The statistical precision of such an experiment is relatively low. But this lack of precision is the price for an unusually broad focus: the study surveys the whole spectrum of cases that come before the American jury. In this sense, the research design may fittingly be called a survey experiment.

From these data it was possible to determine how often judge and jury agree or disagree, when they disagree, and one could trace and count the reasons for their disagreement since through some fifty-odd questions most detailed information became available on every case.

Hans Zeisel

The first part of this study on the role of the jury in criminal trials has just been published.[28] Table 4-5 reproduces one of its basic findings.

TABLE 4-5. *Agreement and Disagreement between Jury and Judge in Criminal Trials*

		Jury		
		Acquits	Convicts	Hangs
Judge	Acquits	▨ 13	2	1
	Convicts	17	▨ 63	4

TOTAL 100%
NUMBER OF TRIALS (3,576)

▨ = Agreement

Judge and jury agree in (13 + 63 =) 76 per cent of all cases; and of the 24 per cent disagreement cases, the jury is found on the defendant's side in (17 + 4 =) 21 per cent and on the prosecutor's side in (1 + 2 =) 3 per cent of these cases.

The evaluation of the jury as an institution hinges, of course, not only on the *extent* of its disagreement with the judge but on the reasons that produce these disagreements, and it is the presentation and analysis of these reasons that form the main body of the study. They range from different sentiments on the law which the jury entertains (in spite of what law the judge may give them) through sentiments concerning the particular defendant, different views on the weight of the evidence, and occasionally to an imbalance between the performance of the prosecutor and defense counsel in the trial; a few times it will even happen that the disagreement arises from a discrepancy between what jury and judge know about the case.

These are the five somewhat abstract categories into which the specific causes of the various individual disagreements were ultimately summarized.

At this early point it is not possible to assay the practical consequences of this study. At a minimum, it will lend focus and precision to an important debate in which both sides so far have never been able to draw on more than anecdotal support.[29]

Hans Zeisel

Secondary Analysis

Two studies of contemporary legal problems are distinguished by their being primarily reanalyses of data collected in normal administrative routine: one concerned the alleged deterrent effect of capital punishment, the other the problem of court congestion and delay.

In the debate over the merits of capital punishment, the abolitionists had for a long time no good answer to the claim that the death penalty helped to deter would-be murderers. The change came after Thorsten Sellin investigated the problem.[30] He compared homicide rates before and after abolition in some jurisdictions; before and after reintroduction in others; and in jurisdictions that have the death penalty with adjacent jurisdictions that have abolished it. His data made one point clear: whatever other merits the death penalty may have, it has no traceable deterrent effect.

Sellin's data have been quoted wherever the issue is being argued. To what extent they have been an effective cause of abolition is nevertheless difficult to say. Probably they have not been a major cause; growing revulsion from deliberate killing and the actual or near execution of an innocently convicted man have nearly always provided the major impetus. But Sellin's data have helped to silence, if not to convince, a special opposition.

The study of court congestion opened insights into a problem that is less dramatic but in the long run perhaps more persistent. It is one of the puzzling aspects of our judicial system that the adjudication of civil claims in most of our metropolitan courts is scandalously delayed. In Chicago, for instance, it takes on the average five and a half years from the date a claim is filed until it can be tried before a jury.

In 1957, the University of Chicago Law School published a study of this congestion problem[31] which had this methodological distinction: its more than three hundred pages of measurement and analysis were based almost entirely on data that had become available in the course of routine housekeeping by the courts in their normal administrative business. From these data a number of measurements and parameters were developed that were to acquire some currency in the administration of the courts: a basic formula for measuring delay was developed;[32] a variety of remedies was evaluated, and, as the case may be, rejected, recommended, or suggested for further investigation. Among the latter were the pretrial and the split trial discussed above.

With the ever mounting costs of securing primary data, this harvest from secondary analysis holds a promise for social research generally.

Surveys

Although, as we have seen, the border line between experiment and survey is not a sharp one, it is useful to distinguish the two, especially since, in contrast to the experiment, the primary function of the survey is to provide description. Out of the rapidly growing number of surveys undertaken to give guidance to the lawmakers, only two will be mentioned here. Both deal with most acute legal problems; one with the enforcement of civil-rights legislation, the other with the costs of automobile accidents.

In 1962 Blumrosen and Zeitz began to investigate the operation of New Jersey's antidiscrimination laws and its Civil Rights Commission. Blumrosen examined all cases filed with the commission during the fiscal years 1962–1963, and Zeitz made a survey among Negroes on their attitudes toward enforcement of these laws.[33] The main finding of their study was that "the laws of New Jersey against discrimination were not meaningfully and effectively enforced," partly because the Negroes themselves shied away from individual enforcement and partly because of shortcomings in the commission itself. The study had a number of traceable effects: the legislature substantially increased the commission's budget; the commission itself sharpened some of its policies and much of its mode of operations; and the state's attorney-general would speak of his civil-rights division as having been "a shield" so far but now to become "a sword."[34]

The survey on the costs of automobile accidents is the latest on an issue of long-standing concern to the law and has a distinguished research history. The first major study was published in 1932 under the auspices of the Columbia University Committee for Research in the Social Sciences.[35] Like the latest, the Michigan study, it too was a joint effort of lawyers and social scientists. It surveyed broadly, if somewhat haphazardly, the reparation problem caused by automobile accidents. The evidence was drawn from a variety of sources, but even where survey data were used, no claim was made to precision and completeness. Nevertheless, as in many first approaches, the outlines of the problem and the areas of research emerged with great clarity and thus marked an important beginning. The next step came in 1953 when Professor Adams of the Business School at Temple University studied the financial and legal history of a random sample of one hundred automobile accidents in the City of Philadelphia and thus established the pattern for later efforts.[36] The Philadelphia survey displayed all the glories and some of the inadequacies of an inspired, pioneering, shoestring operation. The present study, undertaken jointly by lawyers and social scientists at the University of Michigan, is the apex of this development.[37]

It covers all individuals killed or injured in automobile accidents that occurred in the State of Michigan during one calendar year and ingeniously combines two samples to represent this universe: one taken from the files of the police, the other from the files of the courts where personal-injury claims are litigated. The difficulty of the yet unfinished court case was elegantly solved by substituting the results reached in a comparable group of earlier long-delayed cases. The major research instrument was a mail questionnaire to the parties concerned or to their heirs, thoughtfully supplemented at critical points with personal interviews, especially with the plaintiffs' lawyers in the cases.

We now have reliable, precise, quantitative information on almost every aspect of the injury-reparation process and hence a sound factual basis for the many debates which are currently raging over that problem area. To be sure, we have this knowledge only for one year and only for the State of Michigan. But, the United States being what it is, one should not be in danger if one generalizes from these findings. If there are doubts, they can be removed by duplication of the study elsewhere.

The survey provides, as any good survey should, information on both details and broad outlines. Roughly one of every hundred Michigan residents suffered some loss in an automobile accident during a year. For over 60 per cent of the persons involved in accidents, the loss was below $500; another 30 per cent suffered losses between $500 and $3,000; and the remaining 2 or 3 per cent suffered losses beyond $70,000.

In terms of all victims or their heirs, 23 per cent received no compensation from any source, 37 per cent received some tort liability settlement, and about half of the victims received some compensation from loss, collision, medical care or life insurance. But in terms of the total dollar amount paid to all victims, almost half of the total damages remained uncompensated. The sources of total compensation were tort liability, 55 per cent; loss liability, 38 per cent; workmen's compensation and social security, 7 per cent. The surprising finding is the great role played by loss insurance.

Finally, there is a group of qualitative surveys with hardly a number in them that nevertheless fulfill an important function. They all aim at illuminating the dark corners of the criminal law process. In theory this process is governed only by the law and the courts, but in practice the great bulk of it is decided by the informal practices of the police and the prosecutor's office. The major effort is the American Bar Foundation's Survey of Criminal Justice, of which Professor Remington is the editor.[38] The first volume in this series is a study of arrest procedures.[39] Its raw material are some two thousand field reports of on-the-scene reports. The unending variety of arrest decisions are presented in a meaningful framework, with just the right amount of detail to be both vivid and allow their classification. A systematic inventory is obtained of the variety of procedures, and thus the essential groundwork is laid for a later quantitative survey.[40]

The Trend

The sharp increase in recent years in the number and quality of social-science investigations of legal institutions was spawned by a number of convergent developments. There was first a jurisprudential movement, the Realists, who, beginning in the twenties, asked that the law in action be explored in contrast to the law on the books. Strangely enough, their aim remained for a long time no more than a battle cry. Only now, a generation later, does it assume substance.[41] The second source was the rapid development of research techniques and a concomitant growth of sociological research in general.[42]

It now appears that we are at the threshold of an era in which the lawmakers will find increasing use for empirical social-science research in the sound expectation that it is bound to alleviate their difficult task: to make good law.

REFERENCES

1. In modern times, William Petty's survey of Ireland, made at the time of the Cromwellian conquest, is perhaps one of the first systematic surveys conducted for the lawmaker. British, and to a much lesser extent also American, legislation has a tradition of relying on systematically collected facts whenever broad legislative issues are at debate. Royal Commissions, Select Committees, and Ad Hoc Committees in the British Empire and Commonwealth, Congressional and other legislative and administrative committees in the United States, have made in their time major contributions to the law and incidentally also the body of social science. See Marie Jahoda, Paul F. Lazarsfeld, Hans Zeisel, *Marienthal: Zur Geschichte der Soziographie*, new ed. (Allensbach: Verlag für Demoskopie, 1960), Appendix.
2. Ironically, the most famous of these social-science footnotes, in the celebrated decision of the United States Supreme Court in *Brown v. Board of Education*, the school desegregation case, refers probably to a not very relevant piece of research. It concerned an experiment designed to prove the evil of segregation. The legal scholars seem to be agreed that it did not influence the Court, which was clearly moved by larger, moral considerations; see, for instance, Edmond Cahn, "Jurisprudence," *New York University Law Review*, XXX (1955), 150. There was even considerable debate about the evidential value of that research. See Kenneth B. Clark, "The Desegregation Cases: Criticism of the Social Scientist's Role," *Villanova Law Review*, V (1960), 224, 236; Ernest van den Haag, "Social Science Testimony in the Desegregation Cases—A Reply to Professor Kenneth Clark," *Villanova Law Review*, VI (1960), 69; A. J. Gregor, "The Law, Social Science, and School Segregation: An Assessment," *Western Reserve Law Review*, XIV (1963), 621-636; Ovid C. Lewis, "Parry and Riposte to Gregor's *The Law, Social Scientist, and School Segregation*," *ibid.*, 637.
3. See Hans Zeisel, "The Uniqueness of Survey Evidence," *Cornell Law Quarterly*, XLV (1960), 322.

4. By way of apology: it is impossible to list all such studies; any offered selection is bound to remain arbitrary.
5. The argument that follows has been developed in more detail in Hans Zeisel, "The New York Expert Testimony Project: Some Reflections on Legal Experiments," *Stanford Law Review,* VIII (1956), 730–749, and in Hans Zeisel, Harry Kalven, Jr., and Bernard Buchholz, *Delay in the Courts* (Boston: Little, Brown, 1959), Chapter 21: "The Case for the Official Experiment."
6. California Board of Correction Monographs, Sacramento.
7. See Daniel Glaser, *The Effectiveness of a Prison and Parole System* (Indianapolis: Bobbs-Merrill, 1964), p. 34.
8. Cf. Zeisel, Kalven, and Buchholz, *op. cit.*, note 5, pp. 143 ff.
9. Maurice Rosenberg, *The Pretrial Conference and Effective Justice* (Columbia University Press, New York and London, 1964).
10. The control group was allowed to have optional pretrial, because it was thought that to simply deprive these litigants of their right to pretrial might engender constitutional difficulties.
11. One may ask why the state did not abolish pretrial altogether. One answer is that the particular experiment answered the issue conclusively only with respect to the option alternative. Moreover, pretrial had deep roots in the state's tradition, and the limited design of the experiment probably offered a welcome pretext for a compromise.
12. The pioneering study was Arthur Beeley, *The Bail System in Chicago* (Chicago: University of Chicago Press, 1927). A survey of the literature can be found in Daniel Freed and Patricia Wald, *Bail in the United States* (Washington, D.C., 1964), pp. 9–21. The attack that brought the reform movement to fruition began with a distinguished study by Professor Caleb Foote of the University of Pennsylvania Law School. On the latest developments see C. E. Ares, A. Rankin, and H. Sturz, "The Manhattan Bail Project: An Interim Report on the Use of Pre-Trial Parole," *New York University Law Review,* XXXVIII (1963), 67. On the long-range implication, see also Caleb Foote, "The Coming Constitutional Crisis in Bail," *University of Pennsylvania Law Review,* CXIII (1965), 959.
13. National Conference on Bail and Criminal Justice, *Interim Report* (Washington, D.C.: 1965), p. xxv.
14. Normally, it is considered essential that in a controlled experiment the experimenter be in direct control of the experimental variable: if two types of fertilizer are to be tested, he must be able to control their assignment to the various plots of land. But in the realm of the law such direct control is rarely feasible. Even in the two experiments discussed so far, the ultimate experimental variable was not under the control of the experimenter: in the bail-bond experiment, the final decision whether or not to release the defendant without bail was up to the judge; and the decision whether or not a pretrial was to be held was, in fact, left to the litigants. Yet because of the prior randomization of the experimental and the control group, these decisions did not invalidate the controlled character of the experiments. For general exposition of this approach, see Irwin Towers, Leo Goodman, and Hans Zeisel, "A Method of Measuring the Effects of Television through Controlled Field Experiments," *Studies in Communication,* IV (1962), 87.
15. In some states, such as California, the court pronounces only the sentence range as provided by the statute; the prison authority makes the final decision.
16. Everson, "The Human Element in Justice," *Journal of Criminal Law and Criminology,* X (1919), 90.
17. Gaudet, "Individual Differences in the Sentencing Tendencies of Judges," *Archives*

of Psychology, XXXII (1938), 5. Edward Green, *Judicial Attitudes in Sentencing* (New York: St. Martin's Press, 1961), based on data from Criminal Court in Pennsylvania, is the only study that claims to have found no such differences, but its statistical analysis is open to criticism.

18. S. Sidney Ulmer, "The Political Variable in the Michigan Supreme Court," *Journal of Public Law*, XI (1962), 352.

19. C. Herman Pritchett, "Divisions of Opinion among Justices of the U. S. Supreme Court," *American Political Science Review*, XXXV (1941), 890; later, *The Roosevelt Court: A Study in Politics and Values, 1937–1947* (New York: Macmillan, 1948).

20. Glendon Schubert has been the prime contributor to the field. He has also edited to valuable anthologies which provide a systematic survey of these studies: *Judicial Decision-Making*, (Glencoe, Ill.: The Free Press, 1963) and *Judicial Behavior* (Chicago: Rand McNally, 1964).

21. This suggestion too was made in Zeisel, Kalven, and Buchholz, *op. cit.*, note 5, p. 99.

22. See Hans Zeisel and Thomas Callahan, "Split Trial and Time-Saving: A Statistical Analysis," *Harvard Law Review*, LXXVI (1963), 1606–1625.

23. One of these many court decisions had a direct reference to the Zeisel-Callahan paper: *Driver v. Phillips et al.*, U.S. District Court E.D. Pa., December 14, 1964, 36 FRD 261 (1964).

24. See Rita James Simon, *The American Jury—The Defense of Insanity* (Boston: Little, Brown, 1967).

25. There were also two minor variations built into the experiment pertaining to the quality of the psychiatric expert testimony and to the informing of the jury as to the consequences of a finding of insanity.

26. Simon, *op. cit.* note 18, Chapter 4.

27. Symposium, "Insanity and the Criminal Law—A Critique of Durham v. United States," *University of Chicago Law Review*, XXII (1955), 317.

28. Harry Kalven, Jr., and Hans Zeisel, *The American Jury* (Boston: Little, Brown, 1966).

29. It is, however, possible to cite one early application of findings from the jury study, albeit to a proposed reform of the English jury. The British government has proposed legislation that would allow majority verdicts of 10:2 and 11:1 votes and remove thereby the requirement of unanimity. It was possible to provide the British Home Office with a set of relevant predictions as to what to expect from such a law, because one of the American states, Oregon, allows the majority verdicts Great Britain plans to introduce. The first prediction was that the number of hung juries (juries in which a mistrial is declared because unanimity cannot be reached) would decrease by about 40 per cent. Since about 5 per cent of all trials end in hung juries, this means a change for two out of every hundred trials. The second piece of relevant information was that the presiding judge is by no means always critical of these juries who remain hung at a 10:2 or 11:1 vote; about one-third of these hung juries are characterized as a result at "which a judge too might have come to." Thus, not all of these hung juries are without merit. Third, the Oregon experience suggested that England must henceforth expect some dissent in about one-fourth of all jury verdicts, an experience that might well be shocking in view of the serious sentence that usually follows a verdict of guilty.

30. Thorsten Sellin, in Royal Commission on Capital Punishment, *Minutes of Evidence* (London: H. M. Stat. Off., 1951), p. 647.

31. Zeisel, Kalven, and Buchholz, *op. cit.* note 5.

32. *Ibid.*, pp. 6 and 43. It has now been adopted as a standard measurement by the Administrative Office of the Illinois Courts.
33. A. B. Blumrosen, "Antidiscrimination Laws in Action in New Jersey; A Law-Sociology Study," with a supplement: L. Zeitz, "Survey of Negro Attitudes toward Law," *Rutgers Law Review*, XIX (1965), 189, 288.
34. From a press release by the Office of the Attorney-General (May 12, 1965).
35. Committee to Study Compensation for Automobile Accidents, *Report* (New York: Columbia University Committee for Research in the Social Sciences, 1932).
36. J. Adams, "Economic Financial Consequences of Personal Injuries Sustained in 1953 Philadelphia Automobile Accidents," Temple University *Economics and Business Bulletin*, VII (1955).
37. Alfred F. Conard *et al.*, *Automobile Accident Costs and Payments: Studies in the Economics of Injury Reparation* (Ann Arbor: University of Michigan Press, 1964).
38. Compare, however, also J. Goldstein, "Police Discretion Not to Invoke the Criminal Process: Low Visibility Decisions in the Administration of Justice," *Yale Law Journal*, LXIX (1960), 543, and J. H. Skolnick, *Justice without Trial* (New York: Wiley, 1966).
39. La Fave, *Arrest: The Decision to Take a Suspect into Custody* (Boston: Little, Brown, 1965).
40. Other volumes in this series: D. J. Newman, *Conviction: The Determination of Guilt or Innocence without Trial* (Boston: Little, Brown, 1966); Miller, *Prosecution* (in preparation); Dawson and Ball, *Sentencing* (in preparation).
41. The two most distinguished names in that group were Karl N. Llewellyn, until his recent death professor at the University of Chicago Law School, and Jerome Frank, judge and author.
42. See Hans Zeisel, "Social Research on the Law: The Ideal and the Practical," in William M. Evan, ed., *Law and Sociology* (Glencoe, Ill.: The Free Press, 1962), p. 124.

CHAPTER ONE: LAW AND EMPIRICAL INQUIRY

I

In this book, when we talk about the methodology of the social sciences, we are not referring merely to specific research techniques such as the construction of questionnaires, the use of statistics, and so on. Instead, we are talking about the much broader topic of the nature of empirical inquiry in the area of social behavior--and, in particular, about social behavior involving the legal system. The theoretical questions we ask and the form in which we ask them, the choices we make among the different ways of gathering information, the use of quantification and scientific reasoning, the interpretation and presentation of our findings--all of these, we think, are a part of the methods of the social sciences and they involve imagination and judgment as well as the application of particular technical skills.

From this point of view, the methods of the social sciences--like the methods of all scientific fields--can be seen as involving two rather different elements. On the one hand, there is a highly creative process which follows no known rules. Here we must rely on the art of discovery, the happy accidents by which we stumble into new ideas, so that we understand more than we did before about man and his behavior. On the other hand, there are the definite, prescribed procedures for testing or demonstrating our hypotheses--the established procedures for separating the true from the untrue. These two elements of scientific method might seem to be rather opposed to one another, but in the best of scientific work they blend to form a curious product which is both rigorous and creative, as indicated in the selection from J. Bronowski's Science and Human Values.

II

Now it is possible to teach people something about the techniques of research, although this book certainly does not pretend to cover more than a small portion of the field. Philosophers of science, however, are well agreed that we really do not know very much about how to teach people to be creative in scientific research, any more than we can teach people how to be creative in art. We can set people on the right path--or on what we think is the right path--we can provide what we think is a hospitable atmosphere, we can teach people how to use the tools, and then we must hope for the best. The individual researcher must decide what to study and how to approach his problem; there is no one to tell him with certainty how to make his decisions so that he will contribute something new to the sum of human knowledge.

There is one thing, however, about which we are fairly sure; Almost all significant research has had its origin in a powerful sense of curiosity, a desire to know when confronted with the unknown. And the unknown has usually--although not always--taken the form of a relatively specific question. Some important discoveries have been made by sheer chance, it is true, the result of the most casual and unplanned observation, as in the case of the development of penicillin. Yet, on the whole, scientific knowledge has grown mainly through efforts to find an answer to a clear question.

There is, of course, an important place in science for exploratory work, for the tentative uncovering of a topic about which little is known and no definite questions are possible. But even this exploratory work is best guided by some idea of the kind of question which will prove to be useful or interesting. In general, a student makes a sad mistake if he thinks that he will "find out" about an area of interest by simply going in and looking around. He is all too likely to come out more or less where he started, regrettably still very much in the dark, and facing the likelihood of the same, aimless, looking around all over again.

When approaching a new topic, when one first becomes curious about an area, the first step should be to read about what is already known and the research that has been done in the past. Exploratory empirical study will probably still be needed--will undoubtedly be needed, for in a certain sense a rather unplanned examination of an issue must be a preliminary step. But it should not be mistaken for the whole of research; and it is most productive if it is undertaken with the goal of developing clear questions kept firmly in mind.

In short, we are arguing for the crucial importance of developing hypotheses, that is, questions and their provisional answers which are then subjected to empirical investigation. *A researcher who says "I am interested in such-and-such an area; I am curious about such-and-such a topic" has only just begun to be curious enough. He must now make an assertion and be determined to find out if his assertion is true or false.*

The nature of this process is excellently illustrated in the following account of Semmelweis' search for the causes of childbed fever. Groping, often mistaken, but with a clear problem to be solved and always with a clear answer to be tested for its validity, the work of Semmelweis is a fine example of what scientific research is all about.

III

What kinds of questions are being asked about the law today for which the social sciences are relevant? The article by Professor Stuart Nagel, in the Introduction, suggested that the range of inquiry is almost as broad as the substantive concerns of the law. But from the point of view of research methodology, it is possible to think of these questions as falling into two main types.

First, there are those questions which are primarily concerned with the description of a single characteristic. For example, we may wish to know how many people in a community are opposed to the use of the death penalty, or how many arrests are made without a warrant in a particular jurisidiction, or the average amount of money damages in a certain kind of automobile tort case. These questions can be said to be logically or theoretically simple, although the research operations needed to answer them may be very difficult indeed. Hopefully, such questions will lead to statements about whether a characteristic is present or absent in a population or a person, or about the amount of a characteristic which is present. * The description of a single characteristic is not to be discounted, for it obviously can have many important uses; but it tells us nothing about why a characteristic exists in the fashion in which it does. It gives us no clue for the control or greater understanding of a phenomenon, no signs pointing to an explanation.

Second, there are those questions which are concerned with the relationships between characteristics--and this is the first step in seeking

---

*Characteristics which vary in amount, such as income or age, are usually termed variables; characteristics which vary by type or category, such as race or sex, are usually termed attributes.

greater understanding, predictive power, or control of a phenomenon. For example, we may be interested in the relationship between city size and the amount of crime, or in the relationship between the income of law firms and the amount of unethical behavior exhibited by its members. Such questions can be called complex, as compared to the simple questions we ask about a single characteristic; and these questions about the relationships between characteristics make up the great bulk of research in the social sciences.

The idea of relationship between characteristics is fundamental to all scientific thought; indeed, it is the basic idea of this book and in one form or another will occupy most of our attention. There are a number of ways to talk about a relationship--we can speak of the association between attributes, the correlation between variables, and so on, as we shall see shortly--but the essential idea is the same, underlying the very structure of our thought and all of our dealings with reality; the things of this world hang together in a stable fashion and the way they hang together can be discovered by scientific investigation.

There is one more important point to be noted. Before the actual work of research can begin, we must take our hypothetical assertions--the questions and provisional answers we wish to look into--and link them to the things of this world, that is, the world of sensory perception. Our assertions, after all, consist of words or verbal symbols. These are often very abstract, referring to such things as "discrimination," "fairness," "motive," "right," "socio-economic status," "ethical behavior," and so on. But what do these terms point to? Can we make them precise, can we make them mean the same thing for different observers? Can we measure them or at least tell whether they are present or absent, high or low, of one type or another? These are the problems we attempt to solve by a careful defining of our concepts and establishing working definitions, and this task is discussed in the selection from Research Methods in Social Relations. It is an essential element of designing research and if it is not done well, our research efforts are certain to founder.

IV

Questions about the law and its operation which are currently investigated with the use of the social sciences tend to be of a rather low order of theoretical sophistication. The field is still very much in the stage of uncovering raw empirical regularities whose interpretation or meaning remains in doubt. And the concepts used are apt to be little more than the common sense categories we employ in grappling with immediate social problems.

To a large extent, this reflects the present stage of development of the social sciences, which are only barely beginning to build a body of well-tested, general theory. And, as a consequence, it is extremely difficult to derive a set of questions or hypotheses from accepted assumptions and principles. A great deal of the research in the social sciences--including research in the area of the law--is really an ad-hoc affair, constructing hypotheses on hunches, and without much regard for the definitions, the procedures, and the findings of previous workers.

But all this is merely to say that our studies of man and his behavior are in their infancy. And our studies can be improved, if we keep trying to state our questions as precisely as we can, explore and analyse the meaning of our terms, and root our problem in the research of other writers.

The last two selections in this chapter are examples of attempts to state a research problem in law and the social sciences. They are written by highly competent scholars and we think the selections are examples of well done work. At the same time, we think these selections can be criticized for a number of inadequacies, which we are sure the authors themselves would be quick to acknowledge. In particular, these two selections, as good as they are, suffer from two very common weaknesses. First, they pay insufficient attention to the more general phenomenon of which they are specific illustrations; and they thus deprive themselves of the thinking, the writing, and the analogies which would be likely to illuminate the problem under examination. Second, these two selections do not really spell out their hypotheses in specific and explicit detail. This is a step, admittedly, which can be carried too far, for as Professors Lazarsfeld and Rosenberg have pointed out, "In any serious piece of scientific work, we do not deal just with a few terms which are linked together by a few simple propositions. We are confronted with long chains of thought directed toward a specific intellectual end. It can hardly be expected that a writer will spell out every little step in his argument.... It is quite easy at this early stage of the social sciences for an over-supply of rigor to end up as a farce."[1] But we still think it is worthwhile for the researcher to state the theory he is using and the hypotheses he has derived as fully as he can. He should, if at all possible, go beyond the assertion that certain characteristics are related and try to say how they are related and why they are related. He then stands a better chance of collecting information which will provide evidence for or against his arguments. To simply collect data about characteristics which are vaguely expected to be related in some fashion to another characteristic in which you are interested has been called the "carpet sweeper" approach. Almost all of us use it, it is true, to some extent; but we should not let it run away with us.

---

[1] Paul F. Lazarsfeld and Morris Rosenberg, (eds.) *The Language of Social Research*, (New York: The Free Press, 1955), p. 496.

HUMANISM AND SCIENCE

We think the student will find the following selection from Science and Human Values useful in that it affords an insight into the humanism which lies at the heart of the scientific process.

"What a scientist does," says Professor Bronowski, "is compounded of two interests: the interest of his time and of his own interest." What this means is that science is not a mechanical process whereby one goes out and "expose[s himself] like a photographic plate" and then develops the images observed. Rather, it is a creative experience whereby the researcher responds to intuitions and ideas concerning the various forces which surround him. In addition, Bronowski tells us that science, traditionally, is socially relevant. Significant scientific discoveries are usually prompted by the pressing questions of the time in which the scientist lives; and thus his references to Faraday, Newton, and Weiner all have in common the fact that, as scientists, they were engaged with pressing problems of their time.

Reprinted from Science and Human Values by J. Bronowski by permission of Julian Messner, A Division of Simon & Schuster, Inc. Copyright 1956, c 1965, by J. Bronowski, pp. 3-18.

Jacob Bronowski, Science and Human Values
(Englewood Cliffs: Prentice-Hall, Inc., 1966) pp. 3-18

ON a fine November day in 1945, late in the afternoon, I was landed on an airstrip in southern Japan. From there a jeep was to take me over the mountains to join a ship which lay in Nagasaki Harbor. I knew nothing of the country or the distance before us. We drove off; dusk fell; the road rose and fell away, the pine woods came down to the road, straggled on and opened again. I did not know that we had left the open country until unexpectedly I heard the ship's loudspeakers broadcasting dance music. Then suddenly I was aware that we were already at the center of damage in Nagasaki. The shadows behind me were the skeletons of the Mitsubishi factory buildings, pushed backwards and sideways as if by a giant hand. What I had thought to be broken rocks was a concrete power house with its roof punched in. I could now make out the outline of two crumpled gasometers; there was a cold furnace festooned with service pipes; otherwise nothing but cockeyed telegraph poles and loops of wire in a bare waste of ashes. I had blundered into this desolate lanscape as instantly as one might wake among the craters of the moon. The moment of recognition when I realized that I was already in Nagasaki is present to me as I write, as vividly as when I lived it. I see the warm night and the meaningless shapes; I can even remember the tune that was coming from the ship. It was a dance tune which had been popular in 1945, and it was called 'Is You Is Or Is You Ain't Ma Baby?'

These essays, which I have called *Science and Human Values,* were born at that moment. For the moment I have recalled was a universal moment; what I met was, almost as abruptly, the experience of mankind. On an evening like that evening, some time in 1945, each of us in his own way learned that his imagination had been dwarfed. We looked up and saw the power of which we had been proud loom over us like the ruins of Nagasaki.

The power of science for good and for evil has troubled other minds than ours. We are not here fumbling with a new dilemma; our subject and our fears are as old as the tool-making civilizations. Men have been killed with weapons before now: what happened at Nagasaki was only more massive (for 40,000 were killed there by a flash which lasted seconds) and more ironical (for the bomb exploded over the main Christian community in Japan). Nothing happened in 1945 except that we changed the scale of our indifference to man; and conscience, in revenge, for an instant became immediate to us. Before this immediacy fades in a sequence of televised atomic tests, let us acknowledge our subject for what it is: civilization face to face with its own implications. The implications are both the industrial slum which Nagasaki was before it was bombed, and the ashy desolation which the bomb made of the slum. And civilization asks of both ruins, 'Is You Is Or Is You Ain't Ma Baby?'

2

The man whom I imagine to be asking this question, wrily with a sense of shame, is not a scientist; he is civilized man. It is of course more usual for each member of civilization to take flight from its consequences by protesting that others have failed him. Those whose education and perhaps tastes have confined them to the humanities protest that the scientists alone are to blame, for plainly no mandarin ever made a bomb or an industry. The scientists say, with equal contempt, that the Greek scholars and the earnest cataloguers of cave paintings do well to wash their hands of blame; but what in fact are they doing to help direct the society whose ills grow more often from inaction than from error?

This absurd division reached its *reductio ad absurdum,* I think, when one of my teachers, G. H. Hardy, justified his great life work on the ground that it could do no one the least harm—or the least good.[1] But Hardy was a mathematician; will humanists really let him opt out of the conspiracy of scientists? Or are scientists in their turn to forgive Hardy because, protest as he might, most of them learned their indispensable mathematics from his books?

There is no comfort in such bickering. When Shelley pictured science as a modern Prometheus who would wake the world to a wonderful dream of Godwin, he was alas too simple. But it is as pointless to read what has happened since as a nightmare. Dream or nightmare, we have to live our experience as it is, and we have to live it awake. We live in a world which is penetrated through and through by science, and which is both whole and real. We cannot turn it into a game simply by taking sides.

And this make-believe game might cost us what we value most: the human content of our lives. The scholar who disdains science may speak in fun, but his fun is not quite a laughing matter. To think of science as a set of special tricks, to see the scientist as the manipulator of outlandish skills—this is the root of the poison mandrake which flourishes rank in the comic strips. There is no more threatening and no more degrading doctrine than the fancy that somehow we may shelve the responsibility for making the decisions of our society by passing it to a few scientists armored with a special magic. This is another dream, the dream of H. G. Wells, in which the tall elegant engineers rule, with perfect benevolence, a humanity which has no business except to be happy. To H. G. Wells, this was a dream of heaven — a modern version of the idle, harp-resounding heaven of other childhood pieties. But in fact it is the picture of a slave society, and should make us shiver whenever we hear a man of sensibility dismiss science as someone else's concern. The world today is made, it is powered by science; and for any man to abdicate an interest in science is to walk with open eyes towards slavery.

My aim in this book is to show that the parts of civilization make a whole: to display the links which give society its coherence and, more, which give it life. In particular, I want to show the place of science in the canons of conduct which it has still to perfect.

This subject falls into three parts. The first is a study of the nature of the scientific activity, and with it of all those imaginative acts of understanding which exercise 'The Creative Mind.' After this it is logical to ask what is the nature of the truth, as we seek it in science and in social life; and to trace the influence which this search for empirical truth has had on conduct. This influence has prompted me to call the second part 'The Habit of Truth.' Last I shall study the conditions for the success of science, and find in them the values of man which science would have had to invent afresh if man had not otherwise known them: the values which make up 'The Sense of Human Dignity.'

This, then, is a high-ranging subject which is not to be held in the narrow limits of a laboratory. It disputes the prejudice of the humanist who takes his science sourly and, equally, the petty view which many scientists take of their own activity and that of others. When men misunderstand their own work, they cannot understand the work of others; so that it is natural that these scientists have been indifferent to the arts. They have been content, with the humanists, to think science mechanical and neutral; they could therefore justify themselves only by the claim that it is practical. By this lame criterion they have of course found poetry and music and painting at least unreal and often meaningless. I challenge all these judgments.

3

There is a likeness between the creative acts of the mind in art and in science. Yet, when a man uses the word science in such a sentence, it may be suspected that he does not mean what the headlines mean by science. Am I about to sidle away to those riddles in the Theory of Numbers which Hardy loved, or to the heady speculations of astrophysicists, in order to make claims for abstract science which have no bearing on its daily practice?

I have no such design. My purpose is to talk about science as it is, practical and theoretical. I define science as the organization of our knowledge in such a way that it commands more of the hidden potential in nature. What I have in mind therefore is both deep and matter of fact; it reaches from the kinetic theory of gases to the telephone and the suspension bridge and medicated toothpaste. It admits no sharp boundary between knowledge and use. There are of course people who like to draw a line between pure and applied science; and oddly, they are often the same people who find art unreal. To them, the word useful is a final arbiter, either for or against a work; and they use this word as if it can mean only what makes a man feel heavier after meals.

There is no sanction for confining the practice of science in this or another way. True, science is full of useful inventions. And its theories have often been made by men whose imagination was directed by the uses to which their age looked. Newton turned naturally to astronomy because it was the subject of his day, and it was so because finding one's way at sea had long been a practical preoccupation of the society into which he was born. It should be added, mischievously, that astronomy also had some standing because it was used very practically to cast horoscopes. (Kepler used it for this purpose; in the Thirty Years' War he cast the horoscope of Wallenstein which wonderfully told his character, and he predicted a universal disaster for 1634 which proved to be the murder of Wallenstein.[2])

In a setting which is more familiar, Faraday worked all his life to link electricity with magnetism because this was the glittering problem of his day; and it was so because his society, like ours, was on the lookout for new sources of power. Consider a more modest example today: the new mathematical methods of automatic control, a subject sometimes called cybernetics, have been developed now because this is a time when communication and control have in effect become forms of power.[3] These inventions have been directed by social needs, and they are useful inventions; yet it was not their usefulness which dominated and set light to the minds of those who made them. Neither Newton nor Faraday, nor yet Norbert Wiener, spent their time in a scramble for patents.

What a scientist does is compounded of two interests: the interest of his time and his own interest. In this his behavior is no different from any other man's. The need of the age gives its shape to scientific progress as a whole. But it is not the need of the age which gives the individual scientist his sense of pleasure and of adventure, and that excitement which keeps him working late into the night when all the useful typists have gone home at five o'clock. He is personally involved in his work, as the poet is in his, and as the artist is in the painting. Paints and painting too must have been made for useful ends; and language was developed, from whatever beginnings, for practical communication. Yet you cannot have a man handle paints or language or the symbolic concepts of physics, you cannot even have him stain a microscope slide, without instantly waking in him a pleasure in the very language, a sense of exploring his own activity. This sense lies at the heart of creation.[4]

4

The sense of personal exploration is as urgent, and as delightful, to the practical scientist as to the theoretical. Those who think otherwise are confusing what is practical with what is humdrum. Good humdrum work without originality is done every day by everyone, theoretical scientists as well as practical, and writers and painters too, as well as truck drivers and bank clerks. Of course the unoriginal work keeps the world going; but it is not therefore the monopoly of practical men. And neither need the practical man be unoriginal. If he is to break out of what has been done before, he must bring to his own tools the same sense of pride and discovery which the poet brings to words. He cannot afford to be less radical in conceiving and less creative in designing a new turbine than a new world system.

And this is why in turn practical discoveries are not made only by practical men. As the world's interest has shifted, since the Industrial Revolution, to the tapping of new springs of power, the theoretical scientist has shifted his interests too. His speculations about energy have been as abstract as once they were about astronomy; and they have been profound now as they were then, because the man loved to think. The Carnot cycle and the dynamo grew equally from this love, and so did nuclear physics and the German V weapons and Kelvin's interest in low temperatures. Man does not invent by following either use or tradition; he does not invent even a new form of communication by calling a conference of communication engineers. Who invented the television set? In any deep sense, it was Clerk Maxwell who foresaw the existence of radio waves, and Heinrich Hertz who proved it, and J. J. Thomson who discovered the electron. This is not said in order to rob any practical man of the invention, but from a sad sense of justice; for neither Maxwell nor Hertz nor J. J. Thomson would take pride in television just now.

Man masters nature not by force but by understanding. This is why science has succeeded where magic failed: because it has looked for no spell to cast over nature. The alchemist and the magician in the Middle Ages thought, and the addict of comic strips is still encouraged to think, that nature must be mastered by a device which outrages her laws. But in four hundred years since the Scientific Revolution we have learned that we gain our ends only *with* the laws of nature; we control her only by understanding her laws. We cannot even bully nature by any insistence that our work shall be designed to give power over her. We must be content that power is the byproduct of understanding. So the Greeks said that Orpheus played the lyre with such sympathy that wild beasts were tamed by the hand on the strings. They did not suggest that he got this gift by setting out to be a lion tamer.

5

What is the insight with which the scientist tries to see into nature? Can it indeed be called either imaginative or creative? To the literary man the question may seem merely silly. He has been taught that science is a large collection of facts; and if this is true, then the only seeing which scientists need do is, he supposes, seeing the facts. He pictures them, the colorless professionals of science, going off to work in the morning into the universe in a neutral, unexposed state. They then expose themselves like a photographic plate. And then in the darkroom or laboratory they develop the image, so that suddenly and startlingly it appears, printed in capital letters, as a new formula for atomic energy.

Men who have read Balzac and Zola are not deceived by the claims of these writers that they do no more than record the facts. The readers of Christopher Isherwood do not take him literally when he writes 'I am a camera.' Yet the same readers solemnly carry with them from their schooldays this foolish picture of the scientist fixing by some mechanical process the facts of nature. I have had of all people a historian tell me that science is a collection of facts, and his voice had not even the ironic rasp of one filing cabinet reproving another.

It seems impossible that this historian had ever studied the beginnings of a scientific discovery. The Scientific Revolution can be held to begin in the year 1543 when there was brought to Copernicus, perhaps on his deathbed, the first printed copy of the book he had finished about a dozen years earlier. The thesis of this book is that the earth moves around the sun. When did Copernicus go out and record this fact with his camera? What appearance in nature prompted his outrageous guess? And in what odd sense is this guess to be called a neutral record of fact?

Less than a hundred years after Copernicus, Kepler published (between 1609 and 1619) the three laws which describe the paths of the planets. The work of Newton and with it most of our mechanics spring from these laws. They have a solid, matter of fact sound. For example, Kepler says that if one squares the year of a planet, one gets a number which is proportional to the cube of its average distance from the sun. Does anyone think that such a law is found by taking enough readings and then squaring and cubing everything in sight? If he does then, as a scientist, he is doomed to a wasted life; he has as little prospect of making a scientific discovery as an electronic brain has.

It was not this way that Copernicus and Kepler thought, or that scientists think today. Copernicus found that the orbits of the planets would look simpler if they were looked at from the sun and not from the earth. But he did not in the first place find this by routine calculation. His first step was a leap of imagination—to lift himself from the earth, and put himself wildly, speculatively into the sun.[5] 'The earth conceives from the sun,' he wrote; and 'the sun rules the family of stars.' We catch in his mind an image, the gesture of the virile man standing in the sun, with arms outstretched, overlooking the planets. Perhaps Copernicus took the picture from the drawings of the youth with outstretched arms which the Renaissance teachers put into their books on the proportions of the body. Perhaps he had seen Leonardo's drawings of his loved pupil Salai. I do not know. To me, the gesture of Copernicus, the shining youth looking outward from the sun, is still vivid in a drawing which William Blake in 1780 based on all these: the drawing which is usually called *Glad Day*.[6]

Kepler's mind, we know, was filled with just such fanciful analogies; and we know what they were. Kepler wanted to relate the speeds of the planets to the musical intervals. He tried to fit the five regular solids into their orbits. None of these likenesses worked, and they have been forgotten; yet they have been and they remain the stepping stones of every creative mind. Kepler felt for his laws by way of metaphors, he searched mystically for likenesses with what he knew in every strange corner of nature. And when among these guesses he hit upon his laws, he did not think of their numbers as the balancing of a cosmic bank account, but as a revelation of the unity in all nature. To us, the analogies by which Kepler listened for the movement of the planets in the music of the spheres are farfetched.[7] Yet are they more so than the wild leap by which Rutherford and Bohr in our own century found a model for the atom in, of all places, the planetary system?

6

No scientific theory is a collection of facts. It will not even do to call a theory true or false in the simple sense in which every fact is either so or not so. The Epicureans held that matter is made of atoms two thousand years ago and we are now tempted to say that their theory was true. But if we do so we confuse their notion of matter with our own. John Dalton in 1808 first saw the structure of matter as we do today, and what he took from the ancients was not their theory but something richer, their image: the atom. Much of what was in Dalton's mind was as vague as the Greek notion, and quite as mistaken. But he suddenly gave life to the new facts of chemistry and the ancient theory together, by fusing them to give what neither had: a coherent picture of how matter is linked and built up from different kinds of atoms. The act of fusion is the creative act.

All science is the search for unity in hidden likenesses. The search may be on a grand scale, as in the modern theories which try to link the fields of gravitation and electromagnetism. But we do not need to be browbeaten by the scale of science. There are discoveries to be made by snatching a small likeness from the air too, if it is bold enough. In 1935 the Japanese physicist Hideki Yukawa wrote a paper which can still give heart to a young scientist. He took as his starting point the known fact that waves of light can sometimes behave as if they were separate pellets. From this he reasoned that the forces which hold the nucleus of an atom together might sometimes also be observed as if they were solid pellets. A schoolboy can see how thin Yukawa's analogy is, and his teacher would be severe with it. Yet Yukawa without a blush calculated the mass of the pellet he expected to see, and waited. He was right; his meson was found, and a range of other mesons, neither the existence nor the nature of which had been suspected before. The likeness had borne fruit.

The scientist looks for order in the appearances of nature by exploring such likenesses. For order does not display itself of itself; if it can be said to be there at all, it is not there for the mere looking. There is no way of pointing a finger or a camera at it; order must be discovered and, in a deep sense, it must be created. What we see, as we see it, is mere disorder.

This point has been put trenchantly in a fable by Karl Popper. Suppose that someone wished to give his whole life to science. Suppose that he therefore sat down, pencil in hand, and for the next twenty, thirty, forty years recorded in notebook after notebook everything that he could observe. He may be supposed to leave out nothing: today's humidity, the racing results, the level of cosmic radiation and the stockmarket prices and the look of Mars, all would be there. He would have compiled the most careful record of nature that has ever been made; and, dying in the calm certainty of a life well spent, he would of course leave his notebooks to the Royal Society. Would the Royal Society thank him for the treasure of a lifetime of observation? It would not. The Royal Society would treat his notebooks exactly as the English bishops have treated Joanna Southcott's box. It would refuse to open them at all, because it would know without looking that the notebooks contain only a jumble of disorderly and meaningless items.

7

Science finds order and meaning in our experience, and sets about this in quite a different way. It sets about it as Newton did in the story which he himself told in his old age, and of which the schoolbooks give only a caricature. In the year 1665, when Newton was twenty-two, the plague broke out in southern England, and the University of Cambridge was closed. Newton therefore spent the next eighteen months at home, removed from traditional learning, at a time when he was impatient for knowledge and, in his own phrase, 'I was in the prime of my age for invention.' In this eager, boyish mood, sitting one day in the garden of his widowed mother, he saw an apple fall. So far the books have the story right; we think we even know the kind of apple; tradition has it that it was a Flower of Kent. But now they miss the crux of the story. For what struck the young Newton at the sight was not the thought that the apple must be drawn to the earth by gravity; that conception was older than Newton. What struck him was the conjecture that the same force of gravity, which reaches to the top of the tree, might go on reaching out beyond the earth and its air, endlessly into space. Gravity might reach the moon: this was Newton's new thought; and it might be gravity which holds the moon in her orbit. There and then he calculated what force from the earth (falling off as the square of the distance) would hold the moon, and compared it with the known force of gravity at tree height. The forces agreed; Newton says laconically, 'I found them answer pretty nearly.' Yet they agreed only nearly: the likeness and the approximation go together, for no likeness is exact. In Newton's sentence modern science is full grown.

It grows from a comparison. It has seized a likeness between two unlike appearances; for the apple in the summer garden and the grave moon overhead are surely as unlike in their movements as two things can be. Newton traced in them two expressions of a single concept, gravitation: and the concept (and the unity) are in that sense his free creation. The progress of science is the discovery at each step of a new order which gives unity to what had long seemed unlike. Faraday did this when he closed the link between electricity and magnetism. Clerk Maxwell did it when he linked both with light. Einstein linked time with space, mass with energy, and the path of light past the sun with the flight of a bullet; and spent his dying years in trying to add to these likenesses another, which would find a single imaginative order between the equations of Clerk Maxwell and his own geometry of gravitation.

8

When Coleridge tried to define beauty, he returned always to one deep thought: beauty, he said, is 'unity in variety.'[8] Science is nothing else than the search to discover unity in the wild variety of nature—or more exactly, in the variety of our experience. Poetry, painting, the arts are the same search, in Coleridge's phrase, for unity in variety. Each in its own way looks for likenesses under the variety of human experience. What is a poetic image but the seizing and the exploration of a hidden likeness, in holding together two parts of a comparison which are to give depth each to the other? When Romeo finds Juliet in the tomb, and thinks her dead, he uses in his heartbreaking speech the words,

> Death that hath suckt the honey of thy breath.

The critic can only haltingly take to pieces the single shock which this image carries. The young Shakespeare admired Marlowe, and Marlowe's Faustus had said of the ghostly kiss of Helen of Troy that it sucked forth his soul. But that is a pale image; what Shakespeare has done is to fire it with the single word honey. Death is a bee at the lips of Juliet, and the bee is an insect that stings; the sting of death was a commonplace phrase when Shakespeare wrote. The sting is there, under the image; Shakespeare has packed it into the word honey; but the very word rides powerfully over its own undertones. Death is a bee that stings other people, but it comes to Juliet as if she were a flower; this is the moving thought under the instant image. The creative mind speaks in such thoughts.

The poetic image here is also, and accidentally, heightened by the tenderness which town dwellers now feel for country ways. But it need not be; there are likenesses to conjure with, and images as powerful, within the man-made world. The poems of Alexander Pope belong to this world. They are not countrified, and therefore readers today find them unemotional and often artificial. Let me then quote Pope: here he is in a formal satire face to face, towards the end of his life, with his own gifts. In eight lines he looks poignantly forward towards death and back to the laborious years which made him famous.

> Years foll'wing Years, steal something ev'ry day,
> At last they steal us from our selves away;
> In one our Frolicks, one Amusements end,
> In one a Mistress drops, in one a Friend:
> This subtle Thief of Life, this paltry Time,
> What will it leave me, if it snatch my Rhime?
> If ev'ry Wheel of that unweary'd Mill
> That turn'd ten thousand Verses, now stands still.

The human mind had been compared to what the eighteenth century called a mill, that is to a machine, before; Pope's own idol Bolingbroke had compared it to a clockwork. In these lines the likeness goes deeper, for Pope is thinking of the ten thousand Verses which he had translated from Homer: what he says is sad and just at the same time, because this really had been a mechanical and at times a grinding task.[9] Yet the clockwork is present in the image too; when the wheels stand still, time for Pope will stand still for ever; we feel that we already hear, over the horizon, Faust's defiant reply to Mephistopheles, which Goethe had not yet written —'let the clock strike and stop, let the hand fall, and time be at an end.'

> Werd ich zum Augenblicke sagen:
> Verweile doch! du bist so schön!

Dann magst du mich in Fesseln schlagen,
Dann will ich gern zugrunde gehn!
Dann mag die Totenglocke schallen,
Dann bist du deines Dienstes frei,
Die Uhr mag stehn, der Zeiger fallen,
Es sei die Zeit für mich vorbei! [10]

I have quoted Pope and Goethe because their metaphor here is not poetic; it is rather a hand reaching straight into experience and arranging it with new meaning. Metaphors of this kind need not always be written in words. The most powerful of them all is simply the presence of King Lear and his Fool in the hovel of a man who is shamming madness, while lightning rages outside. Or let me quote another clash of two conceptions of life, from a modern poet. In his later poems W. B. Yeats was troubled by the feeling that in shutting himself up to write, he was missing the active pleasures of life; and yet it seemed to him certain that the man who lives for these pleasures will leave no lasting work behind him. He said this at times very simply, too:

The intellect of man is forced to choose
Perfection of the life, or of the work.

This problem, whether a man fulfills himself in work or in play, is of course more common than Yeats allowed; and it may be more commonplace. But it is given breadth and force by the images in which Yeats pondered it.

Get all the gold and silver that you can,
Satisfy ambition, or animate
The trivial days and ram them with the sun,
And yet upon these maxims meditate:
All women dote upon an idle man
Although their children need a rich estate;
No man has ever lived that had enough
Of children's gratitude or woman's love.[11]

The love of women, the gratitude of children: the images fix two philosophies as nothing else can. They are tools of creative thought, as coherent and as exact as the conceptual images with which science works: as time and space, or as the proton and the neutron.

9

The discoveries of science, the works of art are explorations — more, are explosions, of a hidden likeness. The discoverer or the artist presents in them two aspects of nature and fuses them into one. This is the act of creation, in which an original thought is born, and it is the same act in original science and original art. But it is not therefore the monopoly of the man who wrote the poem or who made the discovery. On the contrary, I believe this view of the creative act to be right because it alone gives a meaning to the act of appreciation. The poem or the discovery exists in two moments of vision: the moment of appreciation as much as that of creation; for the appreciator must see the movement, wake to the echo which was started in the creation of the work. In the moment of appreciation we live again the moment when the creator saw and held the hidden likeness. When a simile takes us aback and persuades us together, when we find a juxtaposition in a picture both odd and intriguing, when a theory is at once fresh and convincing, we do not merely nod over someone else's work. We re-enact the creative act, and we ourselves make the discovery again. At bottom, there is no unifying likeness there until we too have seized it, we too have made it for ourselves.

How slipshod by comparison is the notion that either art or science sets out to copy nature. If the task of the painter were to copy for men what they see, the critic could make only a single judgment: either that the copy is right or that it is wrong. And if science were a copy of fact, then every theory would be either right or wrong, and would be so for ever. There would be nothing left for us to say but this is so, or is not so. No one who has read a page by a good critic or a speculative scientist can ever again think that this barren choice of yes or no is all that the mind offers.

Reality is not an exhibit for man's inspection, labelled 'Do not touch.' There are no appearances to be photographed, no experiences to be copied, in which we do not take part. Science, like art, is not a copy of nature but a re-creation of her. We re-make nature by the act of discovery, in the poem or in the theorem. And the great poem and the deep theorem are new to every reader, and yet are his own experiences, because he himself re-creates them. They are the marks of unity in variety; and in the instant when the mind seizes this for itself, in art or in science, the heart misses a beat.

NOTES

1. This is the argument in Hardy's delightful small book *A Mathematician's Apology* (Cambridge, 1940). The title of the book and its date, soon after the outbreak of war, suggest that it was prompted by the same distress at the visible misuse of science that has prompted my book. Hardy was a great mathematician, and he was also (under his restrained manner) a man of passionate human and social sympathies.

2. Wallenstein, the great Catholic general in this religious war, was born at the conjunction of Jupiter and Saturn. There is a conjunction of these two planets when they appear to change places repeatedly: in the language of astrology, when they play. This rare conjunction has foreshadowed a number of historic events, for historic events are always plentiful. It occurred about six years before the birth of Christ, and also before the Protestant king William of Orange came to England in the Glorious Revolution of 1688. The same conjunction of Jupiter and Saturn occurred in 1940, and I will quote a poem which I wrote about it in that year.

Jupiter and Saturn played.
The age was broken and re-made.
A rocket rose from Bethlehem.
Christ marched with the Orangemen;
Till, diving, the exploding light
Struck today, and charred it white.

The rocket roars and plunges out.
Saturn and Jupiter turn about.
No child again shall put to shame
The gunsights trained on Bethlehem;
While ice-cap, omen, march to birth
The orbit of the screaming earth.

It was in observing another rare astrological event, the triple conjunction of Jupiter, Saturn and Mars, that Joannes Brunowski discovered, and informed Kepler of, the famous supernova of 1604.

3. At the time of the Scientific Revolution in the sixteenth century, and for two centuries after it, most self-made men got their wealth by trade (in which I include the support of trade by insurance and banking), and often by oversea trade. As *The Merchant of Venice* reminds us, this is how the great fortunes in North Italy, in Holland and in England were made. It was therefore natural that science in these two centuries was agog with problems of trade, and particularly of navigation. The Industrial Revolution in the eighteenth century shifted the source of wealth from trade to manufacture; and manufacture has needed more and more mechanical energy. Science has therefore been preoccupied in the last two centuries with problems which center on energy—practical problems from the heat engine to the electromagnetic field, and theoretical problems from thermodynamics to atomic structure. Now that we are in sight of having as much energy as we can need, the interest of scientists is moving from the generation of energy to its control, and particularly to the automatic control of power processes, whose tools are the valve, the semi-conductor and the computer. A characteristic invention of the Scientific Revolution was the telescope, of which Galileo heard from Holland, and which he presented to the Doge after a demonstration in the port of Venice in the presence of the Senate in 1609. The characteristic invention of the Industrial Revolution was the power machine which does the routine work of the human muscle. The characteristic invention of the second Industrial Revolution through which we are passing is the control mechanism which does the routine work of the human brain.

4. As an example, consider the practice of mathematics. Mathematics is in the first place a language in which we discuss those parts of the real world which can be described by numbers or by similar relations of order. But with the workaday business of translating the facts into this language there naturally goes, in those who are good at it, a pleasure in the activity itself.

They find the language richer than its bare content; what is translated comes to mean less to them than the logic and the style of saying it; and from these overtones grows mathematics as a literature in its own right. Mathematics in this sense, pure mathematics, is a form of poetry, which has the same relation to the prose of practical mathematics as poetry has to prose in any other language. This element of poetry, the delight in exploring the medium for its own sake, is an essential ingredient in the creative process.

5. This has now been admirably documented by Thomas S. Kuhn in *The Copernican Revolution* (Harvard, 1957). As he shows, from the Neoplatonist elements in the new humanism 'some Renaissance scientists, like Copernicus, Galileo, and Kepler, seem to have drawn two decidedly un-Aristotelian ideas: a new belief in the possibility and importance of discovering simple arithmetic and geometric regularities in nature, and a new view of the sun as the source of all vital principles and forces in the universe.' Kuhn draws particular attention to the influence of the 'symbolic identification of the sun and God' in the *Liber de Sole* of Marsilio Ficino, a central figure (with Pico della Mirandola, who wrote the famous *De Hominis Dignitate*) in the humanist and Neoplatonist academy of Florence in the fifteenth century. This has been elaborated by A. Koyré in *La révolution astronomique* (Paris, 1961). In 1960 Robert McNulty discovered an eyewitness account of Giordano Bruno's lectures on Copernicus at Oxford in 1583 which shows that Bruno drew heavily on Ficino's *De vita coelitus comparanda;* this is discussed by Frances A. Yates in *Giordano Bruno and the Hermetic Tradition* (London, 1964). The general subject has also been attractively discussed recently by Arthur Koestler in *The Sleepwalkers* (London, 1959), and earlier in Pauli's essay on the mystic images in Kepler's science in *Naturerklärung und Psyche* by C. G. Jung and W. Pauli (Zurich, 1952).

6. The derivation of Blake's drawing from the Renaissance studies, by Leonardo and others, of the Vitruvian proportions and mathematical harmonies of the human figure is also discussed by Sir Kenneth Clark in *The Nude* (London, 1956). It was first remarked by Sir Anthony Blunt in the *Journal of the Warburg Institute* in 1938.

7. The music of the spheres was itself a mathematical conception, which had been invented by Pythagoras in the sixth century B.C. Pythagoras taught that the distances between the heavenly bodies match the lengths of the strings that sound the different musical notes. It was deduced that the spheres that carry the heavenly bodies make music as they turn.

8. In one of the places in which Coleridge put forward this definition, the essays *On the Principles of Genial Criticism* (which Coleridge thought 'the best things he had ever written'), he traced it back to Pythagoras: 'The safest definition, then, of Beauty, as well as the oldest, is that of Pythagoras: THE REDUCTION OF MANY TO ONE.'

9. Pope was near the end of his career, and his friends Gay and Arbuthnot

were already dead, when he published these lines in 1737. (They expand a thought from Horace, and his surviving friend Swift was particularly moved by them.) Twenty-five years earlier, as a young man in *The Rape of the Lock*, Pope had pictured the mill as a happy symbol in the ritual of the coffee-table.

> For lo! the Board with Cups and Spoons is crown'd,
> The Berries crackle, and the Mill turns round.

As the eighteenth century moved on, the image of the mill became more menacing in the minds of poets, until Blake in 1804 wrote of 'dark Satanic Mills.' In part the change kept step with the progress of the Industrial Revolution, which Blake, for example, felt very sensitively. But in the main what the romantic poets feared was the new vision of nature as a machine, which Newton's great reputation had imposed. Blake meant by the Satanic Mills not a factory but the imperturbable cosmic mechanism which was now imagined to drive the planets round their orbits. Blake used the words abstract, Newtonian and Satanic with the same meaning, to describe a machinery that seemed to him opposed to organic life. (So John Constable said of a painting which he despised, 'Such things are marvellous and so is watchmaking.') Goethe, who did original work in biology, also disliked Newton's view of science; like other poets of the time, he felt that it turned the world into a clockwork. Yet at the same time religious apologists like William Paley in his *Evidences of Christianity* were using the same analogy to prove that the world, like a clock, must have an intelligent designer. Thus the symbol of the clockwork, and (as T. S. Ashton has pointed out) a new sense of time in general, were critical in the thought of those who lived through the Industrial Revolution.

10. The greatest satire of the First World War, Karl Kraus's *Die Letzten Tage der Menschheit*, contains a moving echo of these lines, which bears on what I have written in the preceding note. In one poem Kraus describes the machine-made murders of modern war as observed by a man *Mit der Uhr in der Hand*—that is, watch in hand. I quote two verses.

> Dort ist ein Mörser. Ihm entrinnt der arme Mann,
> der ihn erfand. Er schützt sich in dem Graben.
> Weil Zwerge Riesen überwältigt haben,
> seht her, die Uhr die Zeit zum Stehen bringen kann!
>
> Wie viel war's an der Zeit, als jenes jetzt geschah?
> Schlecht sieht das Aug, das giftige Gase beizen.
> Doch hört das Ohr, die Uhr schlug eben dreizehn.
> Unsichtig Wetter kommt, der Untergang ist nah.

The same image of the ticking clockwork haunted me when I visited refugee camps after another war, in 1947; and I wrote,

The voice of God that spoke and struck
Was the cuckoo in the clock.
The exiles in the garden heard
The engine tremble in the bird,
Sobbing throat and iron bill:
Time on his springy wheel stood still.

Time began and time runs down.
The voices in the garden drown.
No God from his machine unhands
The exile with a mouth of sand.
The clockwork cuckoo on the hill,
Abrupt and wheeling, stoops to kill.

11. This verse comes from the poem 'Vacillation,' and I have quoted it as Yeats first printed it, for example in *The Winding Stair and other poems.* In his *Collected Poems* soon after, Yeats left out the word *or* in the second line. No doubt the change improves the meter; but since I am here concerned with the contrast between the two images in Yeats's mind, I have given his original text.

SCIENTIFIC RESEARCH AS A COMMUNITY ENTERPRISE

The authors of the following selection are talking about the process of establishing a scientific conclusion. And the essence of such a process is well summarized by the following statement: "Scientific research is a community enterprise, even though single studies are frequently carried out by individual investigators working alone. Each study rests on earlier ones and provides a basis for future ones. The more links that can be established between a given study and other studies or a body of theory, the greater the probable contribution." This process of building upon other studies is essential in developing a science of human behavior.

Two aspects of this communal effort are stressed in the reading: (l) the establishment of working definitions are the means by which abstract conceptions are made clear to the scientific community and the research tied to the general scientific literature. While an investigator may feel that everyone knows what his concepts mean, critics may not agree. (2) What the authors mean in stating that a research project should be formulated in relatively abstract terms is that one should be able to identify his research with a general topic in the existing body of scientific literature. By so doing, one not only adds to an existing body of literature but is able to draw on existing generalizations about the subject under examination.

This material has been reproduced from Research Methods in Social Relations by Claire Selltiz, Marie Jahoda, Morton Deutsch, and Stuart W. Cook.
Copyright 1951, (c) 1959 by Holt, Rinehart and Winston, Inc. No part of this material may be reproduced in any way without the written permission of Holt, Rinehart and Winston.
pp. 42-47 Reprinted by permission.

Claire Selltiz, Research Methods in Social Relations
(New York: Holt, Rinehart, and Winston, 1964),
pp. 42-47

8

Establishing Working Definitions[8]

No matter how simple or how elaborate an investigator's formal definition of his concepts, he must find some way of translating them into observable events if he is to carry out any research. It is not possible to study "national status gain" or "loss" as *such*, since these constructs have no direct counterparts in observable events. The investigator must devise some operations that will produce data he is satisfied to accept as an indicator of his concept. This stage of project formulation may require considerable ingenuity, especially if the constructs are far removed from everyday events and if little research using these constructs has been carried out.

---

[8] The reader may notice a similarity between our concept of *working definitions* and the more commonly used term *operational definitions*. We have avoided this more usual term because it carries with it certain philosophical connotations that we do not wish to go into here.

In the case of the foreign student study, although there had been considerable research on status gain and loss in other contexts, no previous study had worked systematically with the concept of "national status gain or loss," and thus no operations that would constitute working definitions of these terms had been devised. Morris and Davidsen had already made a start on their working definition when they decided that their basic measure would be the discrepancy between the student's own rating of his country and his estimate of Americans' ratings. But in what terms should these ratings be made? Could students be given a list of countries and asked simply to arrange them in order of the regard in which they held them? This seemed too ambiguous a task; from their own reactions, the investigators concluded that countries would be arranged differently on the basis of different criteria. They finally settled on three criteria in terms of which countries were to be rated: standard of living, cultural standards, and political standards. On the basis of these ratings, an index of "national status" would be constructed. (It will be obvious from this example that the step of establishing working definitions is very closely related to decisions both about the data-collection instruments to be used and the pattern of analysis. This is another instance of the way in which earlier steps determine later ones, and of the necessity for anticipating later steps in earlier stages.)

8

Working definitions are adequate if the instruments or procedures based on them gather data that constitute satisfactory indicators of the concepts they are intended to represent. Whether this result has been achieved is often a matter of judgment. An investigator may feel that his data provide reasonably good indicators of his concepts; a critic of the study may feel that they do not. It frequently happens that the investigator himself is aware that his data constitute only a very limited reflection of the concept he has in mind, but, especially in the early stages of research on a problem, he may not be able to devise a more satisfactory one. In any case, although the investigator will usually report his findings in terms of his abstract concepts in order to relate them more readily to other research and to theory, he and his readers must keep in mind that what he has actually found is a relationship between two sets of data that are intended to *represent* his concepts. Thus, what Morris and Davidsen actually found was that the discrepancy between two ratings of the home country was related to the answers to certain questions about the United States. Although they may quite properly interpret this as indicating a relationship between national status gain or loss on the part of a foreign student and his attitude toward the host country, one should remember that this interpretation depends on the assumption that the answers to the interview questions are adequate working definitions of the concepts *national status gain, national status loss,* and *attitude toward the host country.*

Relating the Findings to Other Knowledge

Scientific research is a community enterprise, even though single studies are frequently carried out by individual investigators working alone. Each study rests on earlier ones and provides a basis for future ones. The more links that can be established between a given study and other studies or a body of theory, the greater the probable contribution.

There are two major ways of relating a given study to a larger body of knowledge. One, obviously, is to examine the research and the thinking that has already been done on the given research problem or problems related to it, and to plan the study so that it ties in with this existing work at as many points as possible. The second is to formulate the research problem at a level sufficiently abstract so that findings from the study may be related to findings from other studies concerned with the same concepts.

The study of foreign students which we have been considering was formulated at a very high level of abstraction, and thus could readily be related to other studies of status mobility in different situations and with different populations. In studies that take their impetus from some scientific question, it is usually not difficult to formulate the research problem at a useful level of abstraction, since scientific questions, by their very nature, are likely to be stated in general terms. But studies that arise from the need to answer a practical question may remain at such a specific level that they make no real contribution to knowledge unless the investigator takes pains to transpose the question to a higher level of abstraction.

Let us consider the example of an investigator who is asked by an agency concerned with the improvement of intergroup relations to carry out a survey for the purpose of evaluating the effectiveness of a series of cartoons about prejudice. If he sets up the problem in such a way that he cannot generalize beyond the specific cartoons in question, he will fail in both his scientific and his practical assignment. The cartoons he is about to investigate are unique products and to some extent different from all other cartoons in content and form. If he sets the research process in motion in order to determine the effectiveness of these few cartoons, which may be topical today and forgotten tomorrow, he is involved in a task that will have to begin all over again as soon as it is finished. If he discovers, for example, that a specific cartoon attracts and amuses a part of its audience but is misunderstood by the majority, he has learned little that deserves to be classified as scientific knowledge. Nor can such results provide much guidance for the cartoon producer. In order to remove this limitation on his work, the investigator, before he proceeds to data collection, must reformulate his concrete problem in a manner that will ultimately permit him to draw conclusions about the more general aspects of both the cartoon itself and the response of persons exposed to it. In other words, his concern, in this stage of problem formulation, must be with the generalizability of his results. It is not enough to ask whether one particular cartoon is understood. Rather, he must ask: What aspects of the cartoon are understood—and by whom? To be able to answer the first part of the question, he must analyze the general features of the cartoon. He may emerge with categories such as "satire," "caption required for understanding," etc. If he can then demonstrate through his inquiry that this type of cartoon is misunderstood because it is taken literally instead of satirically, he is in a position to advise the artist to experiment either with nonsatirical cartoons or with ways of making the satire clearer.

Only when a problem is formulated in generally meaningful terms can the social scientist hope to transfer to other problems the knowledge gained from the study of a unique event. Generalizing from the study of unique occurrences requires that the research problem be formulated in more abstract terms than might be necessary if one were concerned only with answering a question about the single event. Every event in human life, when regarded in its full concreteness, is, of course, unique. Yet, however rare and atypical, it becomes a legitimate problem for scientific inquiry *if underlying processes that may occur in other unique configurations are specified.* In this sense an earthquake, the bombing of Hiroshima, or the execution of Mussolini by enraged Italians—unique events in every sense—are subject matter for research,[9] provided the problem is formulated in terms that refer to processes potentially observable on other occasions.

To formulate a research problem in this manner permits the repetition of studies under different unique conditions. This process of repetition is known as the *replication* of research. It is essential to the development of confidence in research findings. For example, Morris and Davidsen's finding of a relationship between national status gain or loss and attitude toward the host country is interesting and suggestive, and its congruence with the general theory of status mobility lends credence to the finding. But it can be accepted as generally true only after it has been repeated with other students, attending different universities, studying in countries other than the United States. Such replication of research will in the end show whether the underlying process that the social scientist had in mind accounted for the relationship between the two sets of observed events, or whether it was accounted for by some as yet undiscovered conditions that characterized the specific study. Although it is essential to search for more general processes, the investigator must maintain a careful balance between his attention to the unique configuration and to the general aspects of his observations. To neglect the unique configuration may lead to false and premature generalizations; to neglect the general aspects may lead to failure to develop principles that can be used in understanding situations other than the specific one studied.

---

[9] Of course, a major difficulty in investigating such events is that of managing to be in a position where one can obtain trustworthy information about the event, its initiating conditions, and its consequences.

STARTING RESEARCH WITH AN HYPOTHESIS

Semmelweis' search for the causes of childbed fever, described in the following selection, provides us with an excellent model of how one engages in the scientific process of investigation. Semmelweis was faced with a specific problem: "a large proportion of the women who were delivered of their babies. . . contracted a serious and often fatal illness known as puerperal fever or childbed fever." He began his investigation by reviewing the existing litereature pertaining to the cause of the fever, and then undertook a long process of data collection, testing a series of hypotheses.

Hempel's account of the search for the causes of childbed fever is, however, more than just an illustration of perseverance in science. It is also a good example of the need to use an inductive as well as a deductive approach to inquiry. In discussing these two approaches to data collection (inductive and deductive) Hempel says: "the maxim that data should be gathered without guidance by antecedent hypotheses about the connections among the facts under study is self-defeating . . . On the contrary, tentative hypotheses are needed to give direction to a scientific investigation

. . ."

From Carl G. Hempel, Philosophy of Natural Science, (c) 1966. Reprinted by permission of Prentice-Hall, Inc., Englewood Cliffs, New Jersey, pp. 3-18.

Carl Hempel, Philosophy of Natural Science
(Englewood Cliffs: Prentice-Hall, Inc., 1966), pp. 3-18

A case history as an example

As a simple illustration of some important aspects of scientific inquiry let us consider Semmelweis' work on childbed fever. Ignaz Semmelweis, a physician of Hungarian birth, did this work during the years from 1844 to 1848 at the Vienna General Hospital. As a member of the medical staff of the First Maternity Division in the hospital, Semmelweis was distressed to find that a large proportion of the women who were delivered of their babies in that division contracted a serious and often fatal illness known as puerperal fever or childbed fever. In 1844, as many as 260 out of 3,157 mothers in the First Division, or 8.2 per cent, died of the disease; for 1845, the death rate was 6.8 per cent, and for 1846, it was 11.4 per cent. These figures were all the more alarming because in the adjacent Second Maternity Division of the same hospital, which accommodated almost as many women as the First, the death toll from childbed fever was much lower: 2.3, 2.0, and 2.7 per cent for the same years. In a book that he wrote later on the causation and the prevention of childbed fever, Semmelweis describes his efforts to resolve the dreadful puzzle.[1]

He began by considering various explanations that were current at the time; some of these he rejected out of hand as incompatible with well-established facts; others he subjected to specific tests.

---

[1] The story of Semmelweis' work and of the difficulties he encountered forms a fascinating page in the history of medicine. A detailed account, which includes translations and paraphrases of large portions of Semmelweis' writings, is given in W. J. Sinclair, *Semmelweis: His Life and His Doctrine* (Manchester, England: Manchester University Press, 1909). Brief quoted phrases in this chapter are taken from this work. The highlights of Semmelweis' career are recounted in the first chapter of P. de Kruif, *Men Against Death* (New York: Harcourt, Brace & World, Inc., 1932).

One widely accepted view attributed the ravages of puerperal fever to "epidemic influences", which were vaguely described as "atmospheric-cosmic-telluric changes" spreading over whole districts and causing childbed fever in women in confinement. But how, Semmelweis reasons, could such influences have plagued the First Division for years and yet spared the Second? And how could this view be reconciled with the fact that while the fever was raging in the hospital, hardly a case occurred in the city of Vienna or in its surroundings: a genuine epidemic, such as cholera, would not be so selective. Finally, Semmelweis notes that some of the women admitted to the First Division, living far from the hospital, had been overcome by labor on their way and had given birth in the street: yet despite these adverse conditions, the death rate from childbed fever among these cases of "street birth" was lower than the average for the First Division.

On another view, overcrowding was a cause of mortality in the First Division. But Semmelweis points out that in fact the crowding was heavier in the Second Division, partly as a result of the desperate efforts of patients to avoid assignment to the notorious First Division. He also rejects two similar conjectures that were current, by noting that there were no differences between the two Divisions in regard to diet or general care of the patients.

In 1846, a commission that had been appointed to investigate the matter attributed the prevalence of illness in the First Division to injuries resulting from rough examination by the medical students, all of whom received their obstetrical training in the First Division. Semmelweis notes in refutation of this view that (*a*) the injuries resulting naturally from the process of birth are much more extensive than those that might be caused by rough examination; (*b*) the midwives who received their training in the Second Division examined their patients in much the same manner but without the same ill effects; (*c*) when, in response to the commission's report, the number of medical students was halved and their examinations of the women were reduced to a minimum, the mortality, after a brief decline, rose to higher levels than ever before.

Various psychological explanations were attempted. One of them noted that the First Division was so arranged that a priest bearing the last sacrament to a dying woman had to pass through five wards before reaching the sickroom beyond: the appearance of the priest, preceded by an attendant ringing a bell, was held to have a terrifying and debilitating effect upon the patients in the wards and thus to make them more likely victims of childbed fever. In the Second Division, this adverse factor was absent, since the priest had direct access to the sickroom. Semmelweis decided to test this conjecture. He persuaded the priest to come by a roundabout route and without ringing of the bell, in order to reach the sick chamber silently and unobserved. But the mortality in the First Division did not decrease.

A new idea was suggested to Semmelweis by the observation that in the First Division the women were delivered lying on their backs; in the Second Division, on their sides. Though he thought it unlikely, he decided "like a drowning man clutching at a straw", to test whether this difference in procedure was significant. He introduced the use of the lateral position in the First Division, but again, the mortality remained unaffected.

At last, early in 1847, an accident gave Semmelweis the decisive clue for his solution of the problem. A colleague of his, Kolletschka, received a puncture wound in the finger, from the scalpel of a student with whom he was performing an autopsy, and died after an agonizing illness during which he displayed the same symptoms that Semmelweis had observed in the victims of childbed fever. Although the role of microorganisms in such infections had not yet been recognized at the time, Semmelweis realized that "cadaveric matter" which the student's scalpel had introduced into Kolletschka's blood stream had caused his colleague's fatal illness. And the similarities between the course of Kolletschka's disease and that of the women in his clinic led Semmelweis to the conclusion that his patients had died of the same kind of blood poisoning: he, his colleagues, and the medical students had been the carriers of the infectious material, for he and his associates used to come to the wards directly from performing dissections in the autopsy room, and examine the women in labor after only superficially washing their hands, which often retained a characteristic foul odor.

Again, Semmelweis put his idea to a test. He reasoned that if he were right, then childbed fever could be prevented by chemically destroying the infectious material adhering to the hands. He therefore issued an order requiring all medical students to wash their hands in a solution of chlorinated lime before making an examination. The mortality from childbed fever promptly began to decrease, and for the year 1848 it fell to 1.27 per cent in the First Division, compared to 1.33 in the Second.

In further support of his idea, or of his *hypothesis*, as we will also say, Semmelweis notes that it accounts for the fact that the mortality in the Second Division consistently was so much lower: the patients there were attended by midwives, whose training did not include anatomical instruction by dissection of cadavers.

The hypothesis also explained the lower mortality among "street births": women who arrived with babies in arms were rarely examined after admission and thus had a better chance of escaping infection.

Similarly, the hypothesis accounted for the fact that the victims of childbed fever among the newborn babies were all among those whose mothers had contracted the disease during labor; for then the infection could be transmitted to the baby before birth, through the common bloodstream of mother and child, whereas this was impossible when the mother remained healthy.

Further clinical experiences soon led Semmelweis to broaden his hypothesis. On one occasion, for example, he and his associates, having carefully disinfected their hands, examined first a woman in labor who was suffering from a festering cervical cancer; then they proceeded to examine twelve other women in the same room, after only routine washing without renewed disinfection. Eleven of the twelve patients died of puerperal fever. Semmelweis concluded that childbed fever can be caused not only by cadaveric material, but also by "putrid matter derived from living organisms."

Basic steps in testing a hypothesis

We have seen how, in his search for the cause of childbed fever, Semmelweis examined various hypotheses that had been suggested as possible answers. How such hypotheses are arrived at in the first place is an intriguing question which we will consider later. First, however, let us examine how a hypothesis, once proposed, is tested.

Sometimes, the procedure is quite direct. Consider the conjectures that differences in crowding, or in diet, or in general care account for the difference in mortality between the two divisions. As Semmelweis points out, these conflict with readily observable facts. There are no such differences between the divisions; the hypotheses are therefore rejected as false.

But usually the test will be less simple and straightforward. Take the hypothesis attributing the high mortality in the First Division to the dread evoked by the appearance of the priest with his attendant. The intensity of that dread, and especially its effect upon childbed fever, are not as directly ascertainable as are differences in crowding or in diet, and Semmelweis uses an indirect method of testing. He asks himself: Are there any readily observable effects that should occur if the hypothesis were true? And he reasons: *If* the hypothesis were true, *then* an appropriate change in the priest's procedure should be followed by a decline in fatalities. He checks this implication by a simple experiment and finds it false, and he therefore rejects the hypothesis.

Similarly, to test his conjecture about the position of the women during delivery, he reasons: *If* this conjecture should be true, *then* adoption of the lateral position in the First Division will reduce the mortality. Again, the implication is shown false by his experiment, and the conjecture is discarded.

In the last two cases, the test is based on an argument to the effect that *if* the contemplated hypothesis, say *H*, is true, *then* certain observable events (e.g., decline in mortality) should occur under specified circumstances (e.g., if the priest refrains from walking through the wards, or if the women are delivered in lateral position); or briefly, if *H* is true, then so is *I*, where *I* is a statement describing the observable occurrences to be expected. For convenience, let us say that *I* is inferred from, or implied by, *H*; and let us call *I* a *test implication of the hypothesis H*. (We will later give a more accurate description of the relation between *I* and *H*.)

In our last two examples, experiments show the test implication to be false, and the hypothesis is accordingly rejected. The reasoning that leads to the rejection may be schematized as follows:

If *H* is true, then so is *I*.
2*a*] But (as the evidence shows) *I* is not true.
――――――――――
H is not true.

Any argument of this form, called *modus tollens* in logic,[2] is deductively valid; that is, if its premisses (the sentences above the horizontal line) are true, then its conclusion (the sentence below the horizontal line) is unfailingly true as well. Hence, if the premisses of (2*a*) are properly established, the hypothesis *H* that is being tested must indeed be rejected.

Next, let us consider the case where observation or experiment bears out the test implication *I*. From his hypothesis that childbed fever is blood poisoning produced by cadaveric matter, Semmelweis infers that suitable antiseptic measures will reduce fatalities from the disease. This time, experiment shows the test implication to be true. But this favorable outcome does not conclusively prove the hypothesis true, for the underlying argument would have the form

If *H* is true, then so is *I*.
2*b*] (As the evidence shows) *I* is true.
――――――――――
H is true.

And this mode of reasoning, which is referred to as the *fallacy of affirming the consequent*, is deductively invalid, that is, its conclusion may be false even if its premisses are true.[3] This is in fact illustrated by Semmelweis' own experience. The initial version of his account of childbed fever as a form of blood poisoning presented infection with cadaveric matter essentially as the one and only source of the disease; and he was right in reasoning that if this hypothesis should be true, then destruction of cadaveric particles by antiseptic washing should reduce the mortality. Furthermore, his experiment did show the test implication to be true. Hence, in this case, the premisses of (*2b*) were both true. Yet, his hypothesis was false, for as he later discovered, putrid material from living organisms, too, could produce childbed fever.

---

[2] For details, see another volume in this series: W. Salmon, *Logic*, pp. 24-25.
[3] See Salmon, *Logic*, pp. 27-29.

Thus, the favorable outcome of a test, i.e., the fact that a test implication inferred from a hypothesis is found to be true, does not prove the hypothesis to be true. Even if many implications of a hypothesis have been borne out by careful tests, the hypothesis may still be false. The following argument still commits the fallacy of affirming the consequent:

If H is true, then so are $I_1, I_2, \ldots, I_n$.
2c] (As the evidence shows) $I_1, I_2, \ldots, I_n$ are all true.
H is true.

This, too, can be illustrated by reference to Semmelweis' final hypothesis in its first version. As we noted earlier, his hypothesis also yields the test implications that among cases of street births admitted to the First Division, mortality from puerperal fever should be below the average for the Division, and that infants of mothers who escape the illness do not contract childbed fever; and these implications, too, were borne out by the evidence — even though the first version of the final hypothesis was false.

But the observation that a favorable outcome of however many tests does not afford conclusive proof for a hypothesis should not lead us to think that if we have subjected a hypothesis to a number of tests and all of them have had a favorable outcome, we are no better off than if we had not tested the hypothesis at all. For each of our tests might conceivably have had an unfavorable outcome and might have led to the rejection of the hypothesis. A set of favorable results obtained by testing different test implications, $I_1, I_2, \ldots, I_n$, of a hypothesis, shows that as far as these particular implications are concerned, the hypothesis has been borne out; and while this result does not afford a complete proof of the hypothesis, it provides at least some support, some partial corroboration or confirmation for it. The extent of this support will depend on various aspects of the hypothesis and of the test data. These will be examined in Chapter 4.

Let us now consider another example,[4] which will also bring to our attention some further aspects of scientific inquiry.

---

[4] The reader will find a fuller account of this example in Chap. 4 of J. B. Conant's fascinating book, *Science and Common Sense* (New Haven: Yale University Press, 1951). A letter by Torricelli setting forth his hypothesis and his test of it, and an eyewitness report on the Puy-de-Dôme experiment are reprinted in W. F. Magie, *A Source Book in Physics* (Cambridge: Harvard University Press, 1963), pp. 70-75.

Carl Hempel

As was known at Galileo's time, and probably much earlier, a simple suction pump, which draws water from a well by means of a piston that can be raised in the pump barrel, will lift water no higher than about 34 feet above the surface of the well. Galileo was intrigued by this limitation and suggested an explanation for it, which was, however, unsound. After Galileo's death, his pupil Torricelli advanced a new answer. He argued that the earth is surrounded by a sea of air, which, by reason of its weight exerts pressure upon the surface below, and that this pressure upon the surface of the well forces water up the pump barrel when the piston is raised. The maximum length of 34 feet for the water column in the barrel thus reflects simply the total pressure of the atmosphere upon the surface of the well.

It is evidently impossible to determine by direct inspection or observation whether this account is correct, and Torricelli tested it indirectly. He reasoned that *if* his conjecture were true, *then* the pressure of the atmosphere should also be capable of supporting a proportionately shorter column of mercury; indeed, since the specific gravity of mercury is about 14 times that of water, the length of the mercury column should be about 34/14 feet, or slightly less than 2½ feet. He checked this test implication by means of an ingeniously simple device, which was, in effect, the mercury barometer. The well of water is replaced by an open vessel containing mercury; the barrel of the suction pump is replaced by a glass tube sealed off at one end. The tube is completely filled with mercury and closed by placing the thumb tightly over the open end. It is then inverted, the open end is submerged in the mercury well, and the thumb is withdrawn; whereupon the mercury column in the tube drops until its length is about 30 inches—just as predicted by Torricelli's hypothesis.

A further test implication of that hypothesis was noted by Pascal, who reasoned that if the mercury in Torricelli's barometer is counterbalanced by the pressure of the air above the open mercury well, then its length should decrease with increasing altitude, since the weight of the air overhead becomes smaller. At Pascal's request, this implication was checked by his brother-in-law, Périer, who measured the length of the mercury column in the Torricelli barometer at the foot of the Puy-de-Dôme, a mountain some 4,800 feet high, and then carefully carried the apparatus to the top and repeated the measurement there while a control barometer was left at the bottom under the supervision of an assistant. Périer found the mercury column at the top of the mountain more than three inches shorter than at the bottom, whereas the length of the column in the control barometer had remained unchanged throughout the day.

The role of induction in scientific inquiry

We have considered some scientific investigations in which a problem was tackled by proposing tentative answers in the form of hypotheses that were then tested by deriving from them suitable test implications and checking these by observation or experiment.

But how are suitable hypotheses arrived at in the first place? It is sometimes held that they are inferred from antecedently collected data by means of a procedure called *inductive inference,* as contradistinguished from deductive inference, from which it differs in important respects.

In a deductively valid argument, the conclusion is related to the premisses in such a way that if the premisses are true then the conclusion cannot fail to be true as well. This requirement is satisfied, for example, by any argument of the following general form:

If *p*, then *q*.
It is not the case that *q*.
It is not the case that *p*.

Brief reflection shows that no matter what particular statements may stand at the places marked by the letters '*p*' and '*q*', the conclusion will certainly be true if the premisses are. In fact, our schema represents the argument form called *modus tollens,* to which we referred earlier.

Another type of deductively valid inference is illustrated by this example:

Any sodium salt, when put into the flame of a Bunsen burner, turns the flame yellow.
This piece of rock salt is a sodium salt.
This piece of rock salt, when put into the flame of a Bunsen burner, will turn the flame yellow.

Arguments of the latter kind are often said to lead from the general (here, the premiss about all sodium salts) to the particular (a conclusion about the particular piece of rock salt). Inductive inferences, by contrast, are sometimes described as leading from premisses about particular cases to a conclusion that has the character of a general law or principle. For example, from premisses to the effect that each of the particular samples of various sodium salts that have so far been subjected to the Bunsen flame test did turn the flame yellow, inductive inference supposedly leads to the general conclusion that all sodium salts, when put into the flame of a Bunsen burner, turn the flame yellow. But in this case, the truth of the premisses obviously does *not* guarantee the truth of the conclusion; for even if it is the case that all samples of sodium salts examined so far did turn the Bunsen flame yellow, it remains quite possible that new kinds of sodium salt might yet be found that do not conform to this generalization. Indeed, even some kinds of sodium salt that have already been tested with positive result might conceivably fail to satisfy the generalization under special physical conditions (such as very strong magnetic fields or the like) in which they have not yet been examined. For this reason, the premisses of an inductive inference are often said to imply the conclusion only with more or less high probability, whereas the premisses of a deductive inference imply the conclusion with certainty.

The idea that in scientific inquiry, inductive inference from antecedently collected data leads to appropriate general principles is clearly embodied in the following account of how a scientist would ideally proceed:

> If we try to imagine how a mind of superhuman power and reach, but normal so far as the logical processes of its thought are concerned, . . . would use the scientific method, the process would be as follows: First, all facts would be observed and recorded, *without selection* or *a priori* guess as to their relative importance. Secondly, the observed and recorded facts would be analyzed, compared, and classified, *without hypothesis or postulates* other than those necessarily involved in the logic of thought. Third, from this analysis of the facts generalizations would be inductively drawn as to the relations, classificatory or causal, between them. Fourth, further research would be deductive as well as inductive, employing inferences from previously established generalizations.[5]

This passage distinguishes four stages in an ideal scientific inquiry: (1) observation and recording of all facts, (2) analysis and classification of these facts, (3) inductive derivation of generalizations from them, and (4) further testing of the generalizations. The first two of these stages are specifically assumed not to make use of any guesses or hypotheses as to how the observed facts might be interconnected; this restriction seems to have been imposed in the belief that such preconceived ideas would introduce a bias and would jeopardize the scientific objectivity of the investigation.

But the view expressed in the quoted passage—I will call it *the narrow inductivist conception of scientific inquiry*—is untenable, for several reasons. A brief survey of these can serve to amplify and to supplement our earlier remarks on scientific procedure.

First, a scientific investigation as here envisaged could never get off the ground. Even its first phase could never be carried out, for a collection of *all* the facts would have to await the end of the world, so to speak; and even all the facts *up to now* cannot be collected, since there are an infinite number and variety of them. Are we to examine, for example, all the grains of sand in all the deserts and on all the beaches, and are we to record their shapes, their weights, their chemical composition, their distances from each other, their constantly changing temperature, and their equally changing distance from the center of the moon? Are we to record the floating thoughts that cross our minds in the tedious process? The shapes of the clouds overhead, the changing color of the sky? The construction and the trade name of our writing equipment? Our own life histories and those of our fellow investigators? All these, and untold other things, are, after all, among "all the facts up to now".

---

[5] A. B. Wolfe, "Functional Economics," in *The Trend of Economics*, ed. R. G. Tugwell (New York: Alfred A. Knopf, Inc., 1924), p. 450 (italics are quoted).

Perhaps, then, all that should be required in the first phase is that all the *relevant* facts be collected. But relevant to what? Though the author does not mention this, let us suppose that the inquiry is concerned with a specified *problem*. Should we not then begin by collecting all the facts—or better, all available data—relevant to that problem? This notion still makes no clear sense. Semmelweis sought to solve one specific problem, yet he collected quite different kinds of data at different stages of his inquiry. And rightly so; for what particular sorts of data it is reasonable to collect is not determined by the problem under study, but by a tentative answer to it that the investigator entertains in the form of a conjecture or hypothesis. Given the conjecture that mortality from childbed fever was increased by the terrifying appearance of the priest and his attendant with the death bell, it was relevant to collect data on the consequences of having the priest change his routine; but it would have been totally irrelevant to check what would happen if doctors and students disinfected their hands before examining their patients. With respect to Semmelweis' eventual contamination hypothesis, data of the latter kind were clearly relevant, and those of the former kind totally irrelevant.

Empirical "facts" or findings, therefore, can be qualified as logically relevant or irrelevant only in reference to a given hypothesis, but not in reference to a given problem.

Suppose now that a hypothesis *H* has been advanced as a tentative answer to a research problem: what kinds of data would be relevant to *H*? Our earlier examples suggest an answer: A finding is relevant to *H* if either its occurrence or its nonoccurrence can be inferred from *H*. Take Torricelli's hypothesis, for example. As we saw, Pascal inferred from it that the mercury column in a barometer should grow shorter if the barometer were carried up a mountain. Therefore, any finding to the effect that this did indeed happen in a particular case is relevant to the hypotheses; but so would be the finding that the length of the mercury column had remained unchanged or that it had decreased and then increased during the ascent, for such findings would refute Pascal's test implication and would thus disconfirm Torricelli's hypothesis. Data of the former kind may be called positively, or favorably, relevant to the hypothesis; those of the latter kind negatively, or unfavorably, relevant.

In sum, the maxim that data should be gathered without guidance by antecedent hypotheses about the connections among the facts under study is self-defeating, and it is certainly not followed in scientific inquiry. On the contrary, tentative hypotheses are needed to give direction to a scientific investigation. Such hypotheses determine, among other things, what data should be collected at a given point in a scientific investigation.

It is of interest to note that social scientists trying to check a hypothesis by reference to the vast store of facts recorded by the U.S. Bureau of the Census, or by other data-gathering organizations, sometimes find to their disappointment that the values of some variable that plays a central role in the hypothesis have nowhere been systematically recorded. This remark is not, of course, intended as a criticism of data gathering: those engaged in the process no doubt try to select facts that might prove relevant to future hypotheses; the observation is simply meant to illustrate the impossibility of collecting "all the relevant data" without knowledge of the hypotheses to which the data are to have relevance.

The second stage envisaged in our quoted passage is open to similar criticism. A set of empirical "facts" can be analyzed and classified in many different ways, most of which will be unilluminating for the purposes of a given inquiry. Semmelweis could have classified the women in the maternity wards according to criteria such as age, place of residence, marital status, dietary habits, and so forth; but information on these would have provided no clue to a patient's prospects of becoming a victim of childbed fever. What Semmelweis sought were criteria that would be significantly connected with those prospects; and for this purpose, as he eventually found, it was illuminating to single out those women who were attended by medical personnel with contaminated hands; for it was with this characteristic, or with the corresponding class of patients, that high mortality from childbed fever was associated.

Thus, if a particular way of analyzing and classifying empirical findings is to lead to an explanation of the phenomena concerned, then it must be based on hypotheses about how those phenomena are connected; without such hypotheses, analysis and classification are blind.

Our critical reflections on the first two stages of inquiry as envisaged in the quoted passage also undercut the notion that hypotheses are introduced only in the third stage, by inductive inference from antecedently collected data. But some further remarks on the subject should be added here.

Induction is sometimes conceived as a method that leads, by means of mechanically applicable rules, from observed facts to corresponding general principles. In this case, the rules of inductive inference would provide effective canons of scientific discovery; induction would be a mechanical procedure analogous to the familiar routine for the multiplication of integers, which leads, in a finite number of predetermined and mechanically performable steps, to the corresponding product. Actually, however, no such general and mechanical induction procedure is available at present; otherwise, the much studied problem of the causation of cancer, for example, would hardly have remained unsolved to this day. Nor can the discovery of such a procedure ever be expected. For—to mention one reason—scientific hypotheses and theories are usually couched in terms that do not occur at all in the description of the empirical findings on which they rest, and which they serve to explain. For example, theories about the atomic and subatomic structure of matter contain terms such as 'atom', 'electron', 'proton', 'neutron', 'psi-function', etc.; yet they are based on laboratory findings about the spectra of various gases, tracks in cloud and bubble chambers, quantitative aspects of chemical reactions, and so forth—all of which can be described without the use of those "theoretical terms". Induction rules of the kind here envisaged would therefore have to provide a mechanical routine for constructing, on the basis of the given data, a hypothesis or theory stated in terms of some quite novel concepts, which are nowhere used in the description of the data themselves. Surely, no general mechanical rule of procedure can be expected to achieve this. Could there be a general rule, for example, which, when applied to the data available to Galileo concerning the limited effectiveness of suction pumps, would, by a mechanical routine, produce a hypothesis based on the concept of a sea of air?

To be sure, mechanical procedures for inductively "inferring" a hypothesis on the basis of given data may be specifiable for situations of special, and relatively simple, kinds. For example, if the length of a copper rod has been measured at several different temperatures, the resulting pairs of associated values for temperature and length may be represented by points in a plane coordinate system, and a curve may be drawn through them in accordance with some particular rule of curve fitting. The curve then graphically represents a general quantitative hypothesis that expresses the length of the rod as a specific function of its temperature. But note that this hypothesis contains no novel terms; it is expressible in terms of the concepts of temperature and length, which are used also in describing the data. Moreover, the choice of "associated" values of temperature and length as data already presupposes a guiding hypothesis; namely, that with each value of the temperature, exactly one value of the length of the copper rod is associated, so that its length is indeed a function of its temperature alone. The mechanical curve-fitting routine then serves only to select a particular function as the appropriate one. This point is important; for suppose that instead of a copper rod, we examine a body of nitrogen gas enclosed in a cylindrical container with a movable piston as a lid, and that we measure its volume at several different temperatures. If we were to use this procedure in an effort to obtain from our data a *general* hypothesis representing the volume of the gas as a function of its temperature, we would fail, because the volume of a gas is a function both of its temperature and of the pressure exerted upon it, so that at the same temperature, the given gas may assume different volumes.

Thus, even in these simple cases, the mechanical procedures for the construction of a hypothesis do only part of the job, for they presuppose an antecedent, less specific hypothesis (i.e., that a certain physical variable is a function of one single other variable), which is not obtainable by the same procedure.

There are, then, no generally applicable "rules of induction", by which hypotheses or theories can be mechanically derived or inferred from empirical data. The transition from data to theory requires creative imagination. Scientific hypotheses and theories are not *derived* from observed facts, but *invented* in order to account for them. They constitute guesses at the connections that might obtain between the phenomena under study, at uniformities and patterns that might underlie their occurrence. "Happy guesses" [6] of this kind require great ingenuity, especially if they involve a radical departure from current modes of scientific thinking, as did, for example, the theory of relativity and quantum theory. The inventive effort required in scientific research will benefit from a thorough familiarity with current knowledge in the field. A complete novice will hardly make an important scientific discovery, for the ideas that may occur to him are likely to duplicate what has been tried before or to run afoul of well-established facts or theories of which he is not aware.

Nevertheless, the ways in which fruitful scientific guesses are arrived at are very different from any process of systematic inference. The chemist Kekulé, for example, tells us that he had long been trying unsuccessfully to devise a structural formula for the benzene molecule when, one evening in 1865, he found a solution to his problem while he was dozing in front of his fireplace. Gazing into the flames, he seemed to see atoms dancing in snakelike arrays. Suddenly, one of the snakes formed a ring by seizing hold of its own tail and then whirled mockingly before him. Kekulé awoke in a flash: he had hit upon the now famous and familiar idea of representing the molecular structure of benzene by a hexagonal ring. He spent the rest of the night working out the consequences of this hypothesis.[7]

---

[6] This characterization was given already by William Whewell in his work *The Philosophy of the Inductive Sciences*, 2nd ed. (London: John W. Parker, 1847); II, 41. Whewell also speaks of "invention" as "part of induction" (p. 46). In the same vein, K. Popper refers to scientific hypotheses and theories as "conjectures"; see, for example, the essay "Science: Conjectures and Refutations" in his book, *Conjectures and Refutations* (New York and London: Basic Books, 1962). Indeed, A. B. Wolfe, whose narrowly inductivist conception of ideal scientific procedure was quoted earlier, stresses that "the limited human mind" has to use "a greatly modified procedure", requiring scientific imagination and the selection of data on the basis of some "working hypothesis" (p. 450 of the essay cited in note 5).

This last remark contains an important reminder concerning the objectivity of science. In his endeavor to find a solution to his problem, the scientist may give free rein to his imagination, and the course of his creative thinking may be influenced even by scientifically questionable notions. Kepler's study of planetary motion, for example, was inspired by his interest in a mystical doctrine about numbers and a passion to demonstrate the music of the spheres. Yet, scientific objectivity is safeguarded by the principle that while hypotheses and theories may be freely invented and *proposed* in science, they can be *accepted* into the body of scientific knowledge only if they pass critical scrutiny, which includes in particular the checking of suitable test implications by careful observation or experiment.

Interestingly, imagination and free invention play a similarly important role in those disciplines whose results are validated exclusively by deductive reasoning; for example, in mathematics. For the rules of deductive inference do not afford mechanical rules of discovery, either. As illustrated by our statement of *modus tollens* above, those rules are usually expressed in the form of general schemata, any instance of which is a deductively valid argument. If premisses of the specified kind are given, such a schema does indeed specify a way of proceeding to a logical consequence. But for any set of premisses that may be given, the rules of deductive inference specify an infinity of validly deducible conclusions. Take, for example, one simple rule represented by the following schema:

$$\frac{p}{p \text{ or } q}$$

It tells us, in effect, that from the proposition that p is the case, it follows that p or q is the case, where p and q may be any propositions whatever. The word 'or' is here understood in the "nonexclusive" sense, so that 'p

---

[7] Cf. the quotations from Kekulé's own report in A. Findlay, *A Hundred Years of Chemistry*, 2nd ed. (London: Gerald Duckworth & Co., 1948), p. 37; and W.I.B. Beveridge, *The Art of Scientific Investigation*, 3rd ed. (London: William Heinemann, Ltd., 1957), p. 56.

or *q*' is tantamount to 'either *p* or *q* or both *p* and *q*'. Clearly, if the premiss of an argument of this type is true, then so must be the conclusion; hence, any argument of the specified form is valid. But this one rule alone entitles us to infer infinitely many different consequences from any one premiss. Thus, from 'the Moon has no atmosphere', it authorizes us to infer any statement of the form 'The Moon has no atmosphere, or *q*', where for '*q*' we may write any statement whatsoever, no matter whether it is true or false; for example, 'the Moon's atmosphere is very thin', 'the Moon is uninhabited', 'gold is denser than silver', 'silver is denser than gold', and so forth. (It is interesting and not difficult to prove that infinitely many different statements can be formed in English; each of these may be put in the place of the variable '*q*'.) Other rules of deductive inference add, of course, to the variety of statements derivable from one premiss or set of premisses. Hence, if we are given a set of statements as premisses, the rules of deduction give no direction to our inferential procedures. They do not single out one statement as "the" conclusion to be derived from our premisses, nor do they tell us how to obtain interesting or systematically important conclusions; they provide no mechanical routine, for example, for deriving significant mathematical theorems from given postulates. The discovery of important, fruitful mathematical theorems, like the discovery of important, fruitful theories in empirical science, requires inventive ingenuity; it calls for imaginative, insightful guessing. But again, the interests of scientific objectivity are safeguarded by the demand for an *objective validation* of such conjectures. In mathematics, this means *proof* by deductive derivation from axioms. And when a mathematical proposition has been proposed as a conjecture, its proof or disproof still requires inventiveness and ingenuity, often of a very high caliber; for the rules of deductive inference do not even provide a general mechanical procedure for constructing proofs or disproofs. Their systematic role is rather the modest one of serving as *criteria of soundness for arguments* offered as proofs: an argument will constitute a valid mathematical proof if it proceeds from the axioms to the proposed theorem by a chain of inferential steps each of which is valid according to one of the rules of deductive inference. And to check whether a given argument is a valid proof in this sense is indeed a purely mechanical task.

Scientific knowledge, as we have seen, is not arrived at by applying some inductive inference procedure to antecedently collected data, but rather by what is often called "the method of hypothesis", i.e. by inventing hypotheses as tentative answers to a problem under study, and then subjecting these to empirical test. It will be part of such test to see whether the hypothesis is borne out by whatever relevant findings may have been gathered before its formulation; an acceptable hypothesis will have to fit the available relevant data. Another part of the test will consist in deriving new test implications from the hypothesis and checking these by suitable observations or experiments. As we noted earlier, even extensive testing with entirely favorable results does not establish a hypothesis conclusively, but provides only more or less strong support for it. Hence, while scientific inquiry is certainly not inductive in the narrow sense we have examined in some detail, it may be said to be *inductive in a wider sense,* inasmuch as it involves the acceptance of hypotheses on the basis of data that afford no deductively conclusive evidence for it, but lend it only more or less strong "inductive support", or confirmation. And any "rules of induction" will have to be conceived, in analogy to the rules of deduction, as canons of validation rather than of discovery. Far from generating a hypothesis that accounts for given empirical findings, such rules will presuppose that both the empirical data forming the "premisses" of the "inductive argument" and a tentative hypothesis forming its "conclusion" are *given.* The rules of induction would then state criteria for the soundness of the argument. According to some theories of induction, the rules would determine the strength of the support that the data lend to the hypothesis, and they might express such support in terms of probabilities. In chapters 3 and 4 we will consider various factors that affect the inductive support and the acceptability of scientific hypotheses.

THE NEED FOR ASSERTIONS

The following reading from The American Jury is a good example of socio-legal research undertaken in an inductive fashion. After setting out the various points of view concerning the value of the jury system, Kalven and Zeisel tell us that they will "study the performance of the jury measured against the performance of the judge . . . and that their study will have relevance for both the critics and the defenders of the jury system and will provide fresh material for the jury debate."

We are not adverse to this approach to socio-legal research, but we do feel that Hempel's comment concerning the power of research based on prior theory is applicable to the jury study. He says: "Empirical facts or findings . . . can be qualified as logically relevant or irrelevant only in reference to a given hypothesis, but not in reference to a given problem." If data are to have a powerful impact on understanding a problem, they are best collected in reference to particular assertions concerning the relationships between various characteristics.

From Harry Kalven, Jr. and Hans Zeisel, The American Jury, (Boston: Little, Brown and Company, 1966) pp 3-11. Reprinted by permission.

Harry Kalven, Jr. and Hans Zeisel, The American Jury (Boston: Little, Brown and Co., 1966), pp. 3-11

The Anglo-American jury is a remarkable political institution. We have had it with us for so long that any sense of surprise over its main characteristics has perhaps somewhat dulled. It recruits a group of twelve laymen, chosen at random from the widest population; it convenes them for the purpose of the particular trial; it entrusts them with great official powers of decision; it permits them to carry on deliberations in secret and to report out their final judgment without giving reasons for it; and, after their momentary service to the state has been completed, it orders them to disband and return to private life. The jury thus represents a deep commitment to the use of laymen in the administration of justice, a commitment that finds its analogue in the widespread use of lay judges in the criminal courts of other countries.[1] It opposes the cadre of

---

[1] Lay judges have been a persistent part of the administration of criminal justice since antiquity. During the Middle Ages Europe developed two forms of lay participation in the criminal process, independently both of each other and of the Greco-Roman tradition: on the European continent, the scabini or Schöffen, and in England the jury. See Dawson, A History of Lay Judges (1960).

The Schöffen courts have survived in the mixed tribunals which today are the principal triers of criminal cases in Central, Eastern and Northern Europe. The jury migrated from its English home in two directions. The expanding orbit of the English law took it through what was then the British Empire. The French revolution and its aftermath brought it to Europe, where since Montesquieu's *Esprit des Lois* the jury had been looked at as democracy's way of administering criminal justice. From France the jury spread across Europe and further, first under the direct impact of the revolution and the Napoleonic conquest and, after 1848, more permanently when the second revolutionary wave professional, experienced judges with this transient, ever-changing, ever-inexperienced group of amateurs. The jury is thus by definition an exciting experiment in the conduct of serious human affairs, and it is not surprising that, virtually from its inception, it has been the subject of deep controversy, attracting at once the most extravagant praise and the most harsh criticism.

The jury controversy has recruited some of the great names of political philosophy and the law: Alexander Hamilton, de Tocqueville, Blackstone, Montesquieu, Bentham, Spencer, Livingston, Holmes, Stephen, Corbin, Wigmore, Pound, Sunderland; and, more recently, Frank, Curtis, Green, Wyzanski, Bok, Glanville Williams, Denning, Devlin, and Griswold.[2]

Yet this long tradition of controversy over the jury system

---

carried democratic institutions eastward across the continent as far as Czarist Russia, which instituted jury trial in 1864.

On the whole, the European graft did not take well. See Ch. 2, note 3. But while the jury in recent decades has lost some ground, the principle of lay participation in the criminal process is more firmly established than ever. In the large, only the Near East and Japan and some of the Latin American countries have left the criminal trial exclusively to the learned judiciary. For a general bibliography of the jury in the United States and in other countries, see Pound, Seagle, Jury, in Encyc. Soc. Sci., v. 8, pp. 492-502 (1932).

[2] A complete bibliography on praise and blame of the jury would be formidable; we list a sampling: Livingston, A System of Penal Law for the State of Louisiana, pp. 10 et seq. (1833); Pound, Law in Books and Law in Action, Am. L. Rev., v. 44, p. 12 (1910); Sunderland, Verdicts, General and Special, Yale L.J., v. 29, p. 253 (1920); Wigmore, A Program for the Trial of a Jury Trial, J. Am. Jud. Soc., v. 12, p. 166 (1929); Green, Judge and Jury (1930); Frank, Law and the Modern Mind, Ch. XVI (1930); Frank, Courts on Trial (1949); Curtis, The Trial Judge and the Jury, Vand. L. Rev., v. 5, p. 150 (1952); Wyzanski, A Trial Judge's Freedom and Responsibility, Harv. L. Rev., v. 65, p. 1281 (1952); Devlin, Trial by Jury (1956); Williams, The Proof of Guilt (3d ed. 1963). See also, Broeder, The Functions of the Jury: Facts or Fictions, U. Chi. L. Rev., v. 21, p. 386 (1954). A reasonably complete bibliography on the jury debate was prepared by Professor Dale Broeder for the University of Chicago Jury Project as a staff memorandum; ironically it owes its publication to the Congressional hearings on jury tapping. See Hearing Before the Subcommittee to Investigate the Administration of the Internal Security Act of the Senate Committee on the Judiciary, 84th Cong., 1st Sess., pp. 63-81 (1955).

has produced unsatisfactory debate. Much of the criticism has stemmed from not more than the a priori guess that, since the jury was employing laymen amateurs in what must be a technical and serious business, it could not be a good idea. In comparable fashion, the enthusiasts of the jury have tended to lapse into sentimentality and to equate literally the jury with democracy. Not surprisingly, therefore, the very characteristics which the critics point to as defects, the jury's champions herald as assets.

The wide range of opinions about the jury is easily documented.[3] We begin with a very recent statement by one of the critics. The Dean of the Harvard Law School, in the course of his annual report for 1962-1963, made certain recommendations for improving the administration of justice, among them the abolition of the jury in civil cases. Dean Griswold argued:

> The jury trial at best is the apotheosis of the amateur. Why should anyone think that 12 persons brought in from the street, selected in various ways, for their lack of general ability, should have any special capacity for deciding controversies between persons?[4]

The more exasperated form of criticism is illustrated by the following excerpt from an article in the *American Bar Association Journal* in 1924:

> Too long has the effete and sterile jury system been permitted to tug at the throat of the nation's judiciary as it sinks under the smothering deluge of the obloquy of those it was designed to serve. Too long has ignorance been permitted to sit ensconced in the places of judicial administration where knowledge is so sorely needed. Too long has the lament of the Shakespearean character been echoed, "Justice has fled to brutish beasts and men have lost their reason."[5]

[3] The Jury Project has also conducted an opinion survey among the judiciary on the jury system, from which some findings are set forth in Chapter 37. The Jury Project will present at a future time a complete report of the data.

[4] 1962-1963 Harvard Law School Dean's Report, pp. 5-6.

[5] Sebille, Trial by Jury: An Ineffective Survival, A.B.A.J., v. 10, pp. 53, 55 (1924).

And to add still another unfriendly voice, the distinguished English scholar Glanville Williams, in the Seventh Series of Hamlyn Lectures in 1955 had, among other things, this to say of the jury:

> If one proceeds by the light of reason, there seems to be a formidable weight of argument against the jury system. To begin with, the twelve men and women are chosen haphazard. There is a slight property qualification — too slight to be used as an index of ability, if indeed the mere possession of property can ever be so used; on the other hand, exemption is given to some professional people who would seem to be among the best qualified to serve — clergymen, ministers of religion, lawyers, doctors, dentists, chemists, justices of the peace (as well as all ranks of the armed forces). The subtraction of relatively intelligent classes means that it is an understatement to describe a jury, with Herbert Spencer, as a group of twelve people of average ignorance. There is no guarantee that members of a particular jury may not be quite unusually ignorant, credulous, slow-witted, narrow-minded, biased or temperamental. The danger of this happening is not one that can be removed by some minor procedural adjustment; it is inherent in the English notion of a jury as a body chosen from the general population at random.[6]

The defenders of the jury are equally emphatic. Lord Justice Devlin, an experienced and greatly admired English judge, may speak here for them. In 1956, in the Eighth Hamlyn Lecture Series, he said of the jury:

> Each jury is a little parliament. The jury sense is the parliamentary sense. I cannot see the one dying and the other surviving. The first object of any tyrant in Whitehall would be to make Parliament utterly subservient to his will; and the next to overthrow or diminish trial by jury, for no tyrant could afford to leave a subject's freedom in the hands of twelve of his countrymen. So that trial by jury is more than an instrument of justice and more than one wheel of the constitution: it is the lamp that shows that freedom lives.[7]

---

[6] Williams, The Proof of Guilt, pp. 271-272 (3d ed. 1963).

[7] Devlin, Trial by Jury, p. 164 (1956).

Harry Kalven, Jr. and
Hans Zeisel

Justice Devlin found it appropriate to conclude his lectures on the jury by quoting the famous passage from Blackstone, the words of which, he said, are still "after two centuries as fresh and meaningful as when they were written":

> So that the liberties of England cannot but subsist, so long as this *palladium* remains sacred and inviolate; not only from all open attacks, (which none will be so hardy as to make), but also from all secret machinations, which may sap and undermine it; by introducing new and arbitrary methods of trial, by justices of the peace, commissioners of the revenue, and courts of conscience. And however *convenient* these may appear at first, (as doubtless all arbitrary powers, well executed, are the most *convenient*) yet let it be again remembered, that delays, and little inconveniences in the forms of justice, are the price that all free nations must pay for their liberty in more substantial matters; that these inroads upon this sacred bulwark of the nation are fundamentally opposite to the spirit of our constitution; and that, though begun in trifles, the precedent may gradually increase and spread, to the utter disuse of juries in questions of the most momentous concern.[8]

Thus, after two hundred years, the debate over the jury system, with distinguished participants on both sides, is still going on apace.

This is not the occasion to review the debate systematically. It may be useful, however, to suggest its broad outline. The controversy centers around three large issues. First, there is a series of collateral advantages and disadvantages that are often charged against, or pointed to on behalf of, the jury as an institution. In this realm fall such positive points as that the jury provides an important civic experience for the citizen; that, because of popular participation, the jury makes tolerable the stringency of certain decisions; that, because of its transient personnel, the jury acts as a sort of lightning rod for animosity and suspicion which otherwise might center on the more permanent judge; and that the jury is a guarantor of integrity, since it is said to be more difficult to reach twelve men than one. Against such affirmative claims, serious collateral disadvantages have been urged, chiefly that the jury is expensive; that it contributes to delay in civil litigation; that jury service imposes an unfair tax and social cost on those forced to serve; and that, in general, exposure to jury duty disenchants the citizen and causes him to lose confidence in the administration of justice.

---

[8] Commentaries, v. IV, p. 350 (11th ed. 1791).

✓Second, there is a group of issues that touch directly on the competence of the jury. Here the debate has been fascinating but bitter. On the one hand, it is urged that the judge, as a result of training, discipline, recurrent experience, and superior intelligence, will be better able to understand the law and analyze the facts than laymen, selected from a wide range of intelligence levels, who have no particular experience with matters of this sort, and who have no durable official responsibility. On the other hand, it is argued that twelve heads are inevitably better than one; that the jury as a group has wisdom and strength which need not characterize any of its individual members; that it makes up in common sense and common experience what it may lack in professional training, and that its very inexperience is an asset because it secures a fresh perception of each trial, avoiding the stereotypes said to infect the judicial eye.

✓The third group of issues about the jury goes to what is perhaps the most interesting point. The critics complain that the jury will not follow the law, either because it does not understand it or because it does not like it, and that thus only a very uneven and unequal administration of justice can result from reliance on the jury; indeed, it is said that the jury is likely to produce that government by man, and not by rule of law, against which Anglo-American political tradition is so steadfastly set.

This same flexibility of the jury is offered by its champions as its most endearing and most important characteristic. The jury, it is said, is a remarkable device for insuring that we are governed by the spirit of the law and not by its letter; for insuring that rigidity of any general rule of law can be shaped to justice in the particular case. One is tempted to say that what is one man's equity is another man's anarchy.

From even so brief a summary, it is apparent that there is little chance that the debate over the jury will soon be resolved; it is too threaded with difficult value judgments. For the special purposes of this book, however, three characteristics emerge as salient. First, most praise or blame of the jury can come only by way of the comparison of trial by jury with trial by a judge, the one serious and significant alternative to it.

Thus, throughout the jury controversy there is at least the implicit assumption on both sides that the decisions of the jury will sometimes and to some degree be different from those that would be given by the judge in the same case. Its critics point to these differences as evidence of the jury's fallibility and incompetence; its champions point to these differences as proof of the jury's distinctive function and its strength.

Second, most of the unrest over the jury today is limited to its use in civil trials.[9] It is agreed that the case for the jury in criminal trials is different and much stronger.

Third, while in no small part the jury controversy is clearly in the realm of value judgments and is but a variation on the age-old theme of rule versus equity, nevertheless much of the argument appears to rest on assumptions as to what the facts are — the facts, that is, as to how the jury actually performs. The present book reports an investigation into these facts. It is an effort to provide data on the actual decision-making by juries and judges. Its single purpose is to attempt to answer the question when do trial by judge and trial by jury lead to divergent results.

---

[9] See generally, Kalven, The Dignity of the Civil Jury, Va. L. Rev., v. 50, p. 1055 (1964). See also Green, Juries and Justice — The Jury's Role in Personal Injury Cases, 1962 U. Ill. L. F., p. 152; Joiner, Civil Justice and the Jury (1962). The argument about the jury in civil cases has become much involved with the problem of court congestion. For a recent criticism by a distinguished New York judge, see Desmond, Should It Take 34 Months for a Jury Trial? N.Y. Times, Dec. 8, 1963 (Magazine), p. 29. For an estimate of the impact of the jury on delay, see Zeisel, Kalven, and Buchholz, Delay in the Court, Ch. 6 (1959).

As a matter of both theoretical interest and methodological convenience, we study the performance of the jury measured against the performance of the judge as a baseline. Our material is a massive sample of actual criminal jury trials conducted in the United States in recent years. For each of these trials we have the actual decision of the jury and a communication from the trial judge, telling how he would have disposed of that case had it been tried before him without a jury. In this sense, we have been able to execute the grand experiment of having each case, over the wide universe of contemporary jury business in the criminal law, tried by a jury and also by a judge, thus obtaining matched verdicts for study. The result is a systematic view of how often the jury disagrees with the judge, of the direction of such disagreement, and an assessment of the reasons for it.

From what has been said about the method it is not easy to anticipate the full scope of the study. Trial by judge and trial by jury will emerge not simply as two different modes of adjudication, but as different systems of law. In that sense this is a comparative study of two legal cultures, that of the jury and that of the judge.

We would, however, emphasize that this research operation opens only a limited view on the behavior of the jury. We can answer questions about the jury only in terms of a comparison with the judge. If, for example, we wish to know whether a given factor, such as the defendant's being a cripple, or the skill of counsel, or the commenting on evidence by the judge, has any effect on the jury, we can answer only by measuring what effect such factors have on the jury's propensity to agree or disagree with the judge. But while as a matter of clarity, it is important to emphasize this limitation, it is also important not to make too much of it. It will turn out that through this special window one is able to observe a very considerable amount of jury behavior. For the most part, the impact of a factor on the jury's propensity to agree or disagree with the judge is a good, and perhaps the only practicable, index of its impact on the jury.

While this book then is essentially an empirical study of the jury in operation, we would pause to note its broader implications. It is a contribution to what has often been called realist jurisprudence; it is an effort to find out how the law in operation, as contrasted to the law on the books, is working. There are innumerable other instances of unwritten law in action besides those afforded by the jury, and the study may, therefore, be taken as some indication of what empirical efforts to map out the law in action can yield.[10]

Second, as will become more apparent as we proceed, the book links up with questions about legal decision-making generally, and especially with what the American law has found to be an endlessly fascinating topic: the decision-making of judges. Here again the book can be taken as some indication of what empirical study of the judicial process may yield.

In its concern with measuring jury against judge, the study is, of course, a contribution to the debate over the merits of the jury system. A word of qualification, however, is needed. It is not its purpose to decide whether the jury is a good institution. The purpose has been simply to find out how in fact the jury is performing. The study will have relevance for both the critics and the defenders of the jury system and will provide fresh material for the jury debate. It certainly will not terminate it.

---

[10] During recent years a number of rigorous empirical studies have made their appearance, among them, Conard *et al.*, Automobile Accident Costs and Payments (1964); La Fave, Arrest: The Decision to Take a Suspect into Custody (1965); Rosenberg, The Pretrial Conference and Effective Justice (1964). For its tremendous impact the bail bond experiment of the Vera Foundation deserves mention here: Ares, Rankin, and Sturz, The Manhattan Bail Project: In Interim Report on the Use of Pre-Trial Parole, N.Y.U.L. Rev., v. 38, p. 67 (1963). One might note also the emergence of centers attempting to apply social science inquiries to legal institutions at Columbia, Berkeley, and Wisconsin. For a general survey of these developments see Zeisel, The Law, in Uses of Sociology (Lazarsfeld ed. 1966).

THE POWER OF AN HYPOTHESIS

Carlin's study, Lawyers' Ethics, when contrasted with The American Jury, presents an example of socio-legal research approached from a deductive rather than an inductive perspective. Carlin, unlike Kalven and Zeisel, states an hypothesis and then proceeds to collect and analyse data in relation to his assertion. He says: "We shall be principally concerned with influences on ethical conduct arising from characteristics of the lawyer's clientele, of the courts and agencies of government with which he deals, and of his colleague group, and with the patterning of these influences by the system of social stratification in the bar." The hypothesis is not very specific, it is true, but at least it is there. This is not to say that the testing of theoretically derived hypotheses is the only path to knowledge, as we have noted before. But the former has many advantages in empirical research. Carlin is asserting that there is a relationship between political and social variables and ethical behavior. And he is then led to probe into these issues systematically, which is preferable to a rather haphazard gathering of data.

There is, however, another factor involved in the Carlin study which makes it valuable reading for students of socio-legal research. Carlin's research problem is constructed in such a way as to place it within the general body of literature pertaining to ethical behavior among professional groups. His research problem, as was pointed out in an earlier reading, is formulated "at a level sufficiently abstract so that findings from the study may be related to findings from other studies concerned with the same concepts."

From Jerome E. Carlin, Lawyers' Ethics: A Survey of the New York Bar (New York: Russell Sage Foundation, 1966) pp. 3-9. Reprinted by permission.

Jerome E. Carlin, Lawyers' Ethics: A Survey of the New York Bar (New York: Russell Sage Foundation, 1966), pp.3-9

Members of the legal profession in the United States are bound by a set of ethical rules. These are designed to protect clients from exploitation by lawyers, to promote solidarity and effective working relations among colleagues, and to guard against interference with the proper functioning of courts and agencies of government. In addition, the ethical standards of the bar serve to promote a more basic responsibility:

> The lawyer's highest loyalty is at the same time the most intangible. It is a loyalty that runs not to persons, but to procedures and institutions. The lawyer's role imposes on him a trusteeship for the integrity of those fundamental processes of government and self-government upon which the successful functioning of our society depends.[1]

Frequent ethical violations tend to undermine public confidence in the legal profession, and to weaken the integrity of the administration of justice. Thus, the Preamble to the Canons of Professional Ethics warns:

> In America, where the stability of Courts and all departments of government rests upon the approval of the people, it is peculiarly essential that the system for establishing and dispensing justice be . . . so maintained that the public shall have absolute confidence in the integrity and impartiality of its administration. . . . It cannot be so maintained unless the conduct and motives of our profession are such as to merit the approval of all just men.[2]

Compliance with ethical norms cannot be taken for granted, particularly in view of the nature of lawyer's work. As an advocate asserting or challenging claims of right, as counselor and draftsman fashioning the legal framework of collaborative effort, as negotiator and mediator accommodating interests and manipulating institutional structures, the practicing lawyer is constantly confronted with contending interests and conflicting loyalties. He may not only find himself in situations where his own interests run counter to those of his clients, but where service to clients may be inconsistent with his responsibilities as officer of the law. A critical research task, therefore, is to explore the conditions supporting and impairing the lawyer's capacity to carry out his ethical obligations.

Traditional Approaches to Professional Misconduct

Unethical behavior is most often explained as a product of inadequate training, as a failure on the part of the professional school to instill in students a commitment to professional norms and values. Eliot Freidson has observed:

> Deficient behavior on the part of a professional tends to be seen as the result of being a deficient kind of person, or at least of having been inadequately or improperly "socialized" in professional school. The most common remedy for such behavior is to reform the professional curriculum.[3]

The underlying assumption is that commitment to professional norms and values can be learned in the course of professional training, and that, armed with this firm commitment, the practitioner will be disposed to conform to ethical standards. This position is evident in the following statement regarding the functions of medical education:

> Since numerous kinds of pressures may be exerted upon private practitioners to depart from what they know to be the most appropriate kind of medical care, it becomes functionally imperative that they acquire, in medical school, those values and norms which will

> make them less vulnerable to such deviations. It is in this direct sociological sense that the acquisition of appropriate attitudes and values is as central as the acquisition of knowledge and skills to training for the provision of satisfactory medical care.[4]

Similar views are voiced by commentators on legal education. At the Seton Hall Conference on Professional Responsibility, a law school dean asserted:

> It is necessary for the law school wherever possible to incorporate in the curriculum sufficient scientific ethics to enable a well-meaning lawyer to know what he should do and sufficient incentive to lead him to it.[5]

And a panel discussant at a meeting of the American Bar Association maintained that the way to achieve greater professional responsibility is "to engender in each lawyer a higher sense of dedication to the fullest service of mankind. . . . *It must come from the inside.*"[6]

Although there are disagreements over methods, and doubts about specific programs, there is little wavering in the conviction that the professional school can influence ethical behavior and that this task "is or should be one of the major concerns of professional education."[7] All that is required, according to the author of a recent article in the *Journal of Legal Education,* is "an understanding of the intrapsychic factors that so largely determine the shape of manifest behavior."[8]

Professional misconduct is also seen as a result of inadequacies in the canons of ethics, and in the machinery for their enforcement. The usual solution is to modernize the canons and to increase the effectiveness of disciplinary measures. For example, a special committee has recently been established by the American Bar Association to undertake a "broad re-evaluation of the adequacy and effectiveness of the Canons of Professional Ethics."[9] The committee will "carefully evaluate the extent to which departures from high ethical standards and lapses in strict enforcement are related to the content of the Canons."[10] And it is further noted that "appropriate revisions or additions could contribute significantly to more effective grievance procedure, as well as to increasing the level of voluntary compliance."[11]

Both approaches to the problem of unethical conduct focus on *rules* or *norms* as the principal, if not exclusive, instrument of social control. The assumption is that if the rules were properly defined, effectively internalized, and actively enforced they would be faithfully observed.

This point of view tends to conceive of professional norms as existing independently of the social and organizational context within which they are meant to operate. The focus of attention is on what the professional brings with him to the situation—his norms, values, or commitments—and on the punishment of infractions. Little or no attention is given to the context of action except to emphasize the need for more effective internalization of norms to counteract situational pressures.

Commitment to certain norms and values undoubtedly plays a role, and an important one, in the regulation of professional conduct. It is by no means certain, however, that such commitment can be instilled during professional training; indeed, data presented here and in other studies indicate that this is most unlikely. Nor is it certain that norms and values are the principal determinants of behavior. In fact, the main research problem is to examine the varying significance of norms for behavior within particular social contexts.

Ethical conduct is also undoubtedly affected by formal disciplinary measures. How much of an effect these measures have and the way in which they are brought about raise complex and largely unexplored issues. It seems to be the case in the New York City bar that only a small fraction of violators are officially disciplined. It is possible that if more violators were punished the level of conformity might rise. However, even if enforcement proceedings were carried out against a larger proportion of violators, this would still fail to affect the structural conditions leading to widespread violation of ethical standards. Until these conditions are in some way altered, little can be accomplished by merely increasing the number of punitive actions.

In short, neither internalization nor formal enforcement of norms is sufficient to bring about or fully account for differences in adherence to ethical standards. *The social setting of the lawyer's work*, especially the pattern of opportunities and pressures to which he is exposed in his practice, is equally if not more important.

The Approach of This Study

The perspective adopted in this study is to center attention on how the social organization of the profession affects the ethical behavior of lawyers. We shall be principally concerned with influences on ethical conduct arising from characteristics of the lawyer's clientele, of the courts and agencies of government with which he deals, and of his colleague group, and with the patterning of these influences by the system of social stratification in the bar. We shall also examine lawyers' ethical commitments and the contribution of formal disciplinary measures to the maintenance of professional standards.

The critical importance of situational factors rests in part on the obstacles the lawyer encounters in maintaining his professional independence. Threats to his integrity may arise from the character of the market for legal services, from captivity to clients, and from the contaminating effects of certain courts and agencies.

The market for legal services is highly competitive, particularly in areas of practice calling for low-level, standardized skills. In these areas, the lawyer's work frequently overlaps that of other occupational groups such as realtors and accountants, resulting in intense competition if not open rivalry with these groups. The insecurity of the lawyer's practice is intensified, moreover, by the weak and intermittent demand for legal services among lower- to middle-income individuals. Lawyers who handle these nonrepeating matters are forced to seek continually for new legal business, and often find it necessary to establish illicit connections with runners and other intermediaries. Consequently, lawyers who face these market conditions may be unable to preserve either their professional identity or moral integrity.

Lawyers are also likely to become captives of their clients. This is more often the case for practitioners who provide a continuous and broad range of service to their business clients, and who therefore often become involved in client affairs. Thus, 60 per cent of the lawyers in our sample assist their business clients in obtaining financing; an equal proportion are on the lookout for investment opportunities for these clients; close to half are either officers or board members of client corporations (generally taking an active part in corporate affairs); approximately a third hold stock or have other financial interests in such corporations. Under these conditions lawyers may find it extremely difficult to exercise independent judgment or authority. Moreover, involvement may also increase the lawyer's opportunities to exploit clients and make him more vulnerable to improper client demands.

Finally, the very institutions involved in the administration of justice—the courts and agencies of government—may themselves contribute to undermining the professional integrity of lawyers. In so far as these institutions are subject only to weak professional or bureaucratic controls and are open to political influence, practices are likely to develop which may undermine ethical standards. The inability of the legal profession to make more effective use of the courts for controlling the ethical behavior of lawyers stands in sharp contrast with the increasing use of the hospital for maintaining standards in the medical profession.

THE HOW-TO-DO-IT OF HYPOTHESES CONSTRUCTION

The following selection from Goode and Hatt is a brief discussion of the major points which will aid the researcher in constructing hypotheses. Conceptual clarity, empirical referents, specificity of hypothesis, the relation of hypothesis to a body of theory--all are essential parts of research design.

There are two issues in the selection which are particularly worth noting: (1) "an hypothesis must be separated from a moral preachment . . ."; and, (2) "scientific predictions or hypotheses must avoid the trap of selective evidence by being as definite and specific as possible."

If moral preachments are not carefully separated from factual predictions, the results are all too likely to refer to vague feelings which will connote different things to different people. For example, to use the term "bad" to describe the behavior of parents in relation to their children means that anyone reading the results of the study must interpret what the researcher had in mind when he used the word. If, on the other hand, the research problem is stated in such a way as to lead to hypotheses about the relationship between parental behavior and psychic disturbances in the child, the study takes on an objectivity which can be used by others in carrying out further testing and which exposes the results to critical analysis.

As for specificity in the construction of hypotheses, the need is simple and clear-cut. Not to work in specific terms will often render the assertion non-testable. If the concepts used are too vague or too general, there is always the danger that "any occurrence can be interpreted as a fulfillment."

From William J. Goode and Paul K. Hatt, Methods in Social Research (New York: McGraw-Hill, 1952) pp. 68-73; 170-183. Reprinted by permission.

William J. Goode and Paul K. Hatt, Methods in Social Research (New York: McGraw-Hill, 1952), pp. 68-73.

Let us now look at some criteria for judging hypotheses.

The hypotheses must be conceptually clear. Enough emphasis upon this requirement was made in the preceding chapter to require little further elaboration. It should be repeated, however, that this involves two things. The concepts should be clearly defined, operationally if possible. Moreover, they should be definitions which are commonly accepted and communicable rather than the products of a "private world."

What to do: One simple device for clarifying concepts is to write out a list of the concepts used in the research outline. Then try to define them (*a*) in words, (*b*) in terms of particular operations (index calculations, types of observations, etc.), and (*c*) with reference to other concepts to be found in previous research. Talk over each concept with fellow students and other researchers in the field. It will often be found that supposedly simple concepts contain many meanings. Then it is possible to decide which is the desired referent. For systematic conceptual clarification, perform all the operations suggested in Chapter 5.

Hypotheses should have empirical referents. It has also been previously pointed out that scientific concepts must have an ultimate empirical referent. No usable hypotheses can embody moral judgments. Such statements as "criminals are no worse than businessmen," "women should pursue a career," or "capitalists exploit their workers," are no more usable hypotheses than is the familiar proposition that "pigs are well named because they are so dirty" or the classical question, "How many yards of buttermilk are required to make a pair of breeches for a black bull?" In other words, while a hypothesis may study value judgments, such a goal must be separated from a moral preachment or a plea for acceptance of one's values.

What to do: First, analyze the concepts which express attitudes rather than describing or referring to empirical phenomena. Watch for key words such as "ought," "should," "bad," etc. Then transform the notions into more useful concepts. "Bad parents" is a value term, but the researcher may have a definite description in mind: parents who follow such practices as whimsical and arbitrary authoritarianism, inducing psychic insecurity in the child, failure to give love, etc. "Should" is also a value term, but the student may simply mean, "If women do not pursue a career, we can predict emotional difficulties when the children leave home, or we can predict that the society will not be able to produce as much goods," etc. When, instead, we find that our referent is simply a vague feeling, and we cannot define the operations needed to observe it, we should study the problem further and discover what it is that we really wish to investigate.

William J. Goode and
Paul K. Hatt

The hypotheses must be specific. That is, all the operations and predictions indicated by it should be spelled out. The possibility of actually testing the hypothesis can thus be appraised. Often hypotheses are expressed in such general terms, and with so grandiose a scope, that they are simply not testable. Because of their magnitude, such grand ideas are tempting because they seem impressive and important. It is better for the student to avoid such problems and instead develop his skills upon more tangible notions.

By making all the concepts and operations explicit is meant not only conceptual clarity but a description of any *indexes* to be used. Thus, to hypothesize that the degree of vertical social mobility is decreasing in the United States requires the use of indexes. At present there is no satisfactory operational definition of the status levels which define mobility. Therefore, the hypothesis must include a statement of the indexes which are to be used; that is, political office, occupation, effective income, education, etc.

Such specific formulations have the advantage of assuring that research is practicable and significant, in advance of the expenditure of effort. It furthermore increases the validity of the results, since the broader the terms the easier it is to fall into the trap of using *selective evidence.* The fame of most prophets and fortunetellers lies in their ability to state predictions so that almost any occurrence can be interpreted as a fulfillment. We can express this in almost statistical terms: the more specific the prediction, the smaller the chance that the prediction will actually be borne out as a result of mere accident. Scientific predictions or hypotheses must, then, avoid the trap of selective evidence by being as definite and specific as possible.

What to do: Never be satisfied with a general prediction, if it can be broken into more precise subhypotheses. The general prediction of war is not enough, for example: we must specify time, place, and participants. Predicting the general decline of a civilization is not a hypothesis for testing a theory. Again, we must be able to specify and measure the forces, specify the meaning and time of decline, the population segments involved, etc. Often this can be done by conceptual analysis and the formation of related hypotheses; *e.g.,* we may predict that urbanization is accompanied by a decline in fertility. However, we gain in precision if we attempt to define our indexes of urbanization; specify which segments will be affected, and how much (since in the United States the various ethnic and religious segments are affected differently); specify the amount of fertility decline, and the type (percentage childless, net reproduction rate, etc.). Forming sub-hypotheses (1) clarifies the relationship between the data sought and the conclusions; and (2) makes the specific research task more manageable.

Hypotheses should be related to available techniques. In earlier chapters the point was repeatedly made that theory and method are not opposites. The theorist who does not know what techniques are available to test his hypotheses is in a poor way to formulate usable questions.

This is not to be taken as an absolute injunction against the formulation of hypotheses which at present are too complex to be handled by contemporary technique. It is merely a sensible requirement to apply to any problem in its early stages in order to judge its researchability.

There are some aspects of the impossible hypothesis which may make its formulation worth while. If the problem is significant enough as a possible frame of reference it may be useful whether or not it can be tested at the time. The socioeconomic hypotheses of Marx, for example, were not proved by his data. The necessary techniques were not available either then or now. Nevertheless, Marxian frameworks are an important source of more precise, smaller, verifiable propositions. This is true for much of Émile Durkheim's work on suicide. His related formulations concerning social cohesion have also been useful. The work of both men has been of paramount importance to sociology, even though at the time their larger ideas were not capable of being handled by available techniques.

Furthermore, posing the impossible question may *stimulate* the growth of technique. Certainly some of the impetus toward modern developments in technique has come from criticisms against significant studies which were considered inadequate because of technical limitations. In any serious sociological discussion, research frontiers are continuously challenged by the assertion that various problems "ought" to be investigated even though the investigations are presently impossible.

What to do: Look for research articles on the subject being investigated. Make a list of the various techniques which have been used to measure the factors of importance in the study. If you are unable to locate any discussions of technique, you may find it wiser to do a research *on the necessary research techniques.* You may, instead, decide that this lack of techniques means your problem is too large and general for your present resources.

Some items, such as stratification or race attitudes, have been studied by many techniques. Try to discover why one technique is used in one case and not in another. Note how refinements in technique have been made, and see whether one of these may be more useful for your purposes. Look for criticisms of previous research, so as to understand the weaknesses in the procedures followed.

Again, other problems may have been studied with few attempts at precise measurement. Study the literature to see why this is the case. Ascertain whether some subareas (for example, of religious behavior) may be attacked with techniques used in other areas (for example, attitude measurement, stratification measures, research on choice making, etc.).

The hypothesis should be related to a body of theory. This criterion is one which is often overlooked by the beginning student. He is more likely to select subject matter which is "interesting," without finding out whether the research will really help to refute, qualify, or support any existing theories of social relations. A science, however, can be cumulative only by building on an existing body of fact and theory. It cannot develop if each study is an isolated survey.

Although it is true that the clearest examples of crescive theoretical development are to be found in the physical and biological sciences, the process can also be seen in the social sciences. One such case is the development of a set of generalizations concerning the social character of intelligence. The anthropological investigations at the end of the nineteenth century uncovered the amazing variety of social customs in various societies, while demonstrating conclusively that there were a number of common elements in social life: family systems, religious patterns, an organization of the socialization process, etc.

The French school of sociology, including Lucien Lévy-Bruhl, Émile Durkheim, Marcel Mauss, Henri Hubert, and others, formulated a series of propositions, at the turn of the century, which suggested that the intellectual structure of the human mind is determined by the structure of the society. That is, perception and thought are determined by society, not alone by the anatomical structure of our eyes, ears, and other senses. Modes of thought vary from society to society. Some of these formulations were phrased in an extreme form which need not concern us now, and they were often vague. Nevertheless, the idea was growing that the intelligence of a Polynesian native could not be judged by European standards; his thinking was qualitatively, not merely quantitatively, different.

At the same time, however, better techniques were being evolved for measuring "intelligence," which came to be standardized in the form of scores on various IQ tests. When these were applied to different groups it became clear that the variation in IQ was great; children of Italian immigrants made lower grades on such tests, as did Negroes. Northern Negroes made higher grades than whites from many Southern states. American children of Chinese and Japanese parents made rather high scores. Since it was generally assumed that these tests measured "innate intelligence," these data were sometimes generalized to suggest that certain "racial" groups were by nature inferior and others superior.

However, such conclusions were opposed on rational grounds, and liberal sentiments suggested that they be put to the test. There were, then, two major sets of conclusions, one suggesting that intelligence is in the main determined by social experience, the other suggesting that the IQ is innately determined. To test such opposing generalizations, a research design was needed for testing logical expectations in more specific situations. If, for example, it is true that the intelligence of individuals who are members of "inferior" groups is really determined biologically, then changes in their environments should not change their IQ. If, on the other hand, the social experience is crucial, we should expect that such changes in social experience would result in definite patterns of IQ change.

Further deductions are possible. If identical twins are separated and are placed in radically different social experiences at an early age, we might expect significant differences in IQ. Or, if a group of rural Negro children moves from the poor school and social experience of the South, to the somewhat more stimulating environment of the North, the group averages would be expected to change somewhat. Otto Klineberg, in a classic study, carried out the latter research. He traced Negro children of various ages after they had moved to the North and found that, in general, the earlier the move to the North occurred, the greater the average rise in the IQ. The later the move, the smaller the increase. Even if one assumes that the "better," more able, and more daring adult Negroes made this move, this does not explain the differences by time of movement. Besides, of course, the subjects were children at the time of the migration.[4]

In this research design a particular result was predicted by a series of deductions from a larger set of generalizations. Further, the prediction was actually validated. In justice to the great number of scholars who have been engaged in refining and developing IQ tests, it should be mentioned that other tests and investigations of a similar order have been carried out by many anthropologists, sociologists, and social psychologists. They do not invalidate the notion that IQ is based in part on "innate" abilities, but they do indicate that to a great extent these abilities must be stimulated by certain types of experience in order to achieve high scores on such tests.

From even so sketchy an outline of a theoretical development as the foregoing is, it can be seen that when research is systematically based upon a body of existing theory, a genuine contribution in knowledge is more likely to result. In other words, to be worth doing, a hypothesis must not only be carefully stated, but it should possess theoretical relevance.

What to do: First, of course, cover the literature relating to your subject. If it is impossible to do so, then your hypothesis probably covers too much ground. Second, try to abstract from the literature the way in which various propositions and sets of propositions relate to one another (for example, the literature relating to Sutherland's theory of differential association in criminology, the conditions for maximum morale in factories, or the studies of prediction of marital adjustment). Third, ascertain whether you can deduce any of the propositions, including your own hypothesis, from one another or from a small set of major statements. Fourth, test it by some theoretical model, such as Merton's "Paradigm for Functional Analysis in Sociology" (*Social Theory and Social Structure*, pp. 50–54), to see whether you have left out major propositions and determinants. Fifth, especially compare your own set of related propositions with those of some classic author, such as Weber on bureaucracy or Durkheim on suicide. If you find this task of abstraction difficult, compare instead with the propositions of these men as explained by a systematic interpreter such as Talcott Parsons in his *Structure of Social Action*. What is important is that, whatever the *source* of your hypothesis, it must be *logically* derivable from and based upon a set of related sociological propositions.

---

[4] Otto Klineberg, *Negro Intelligence and Selective Migration* (New York: Columbia University Press, 1935).

SUMMARY

The formulation of the hypothesis is a central step in good research, and it is important to give it a great deal of thought. Because of this significance, we have looked at the hypothesis from several points of view.

1. We have shown why it is so crucial a step to take, and how it functions in a research. It is the question which we put to the empirical world, in such a form that an answer can be obtained.
2. We have also looked at some of the problems which occur when we attempt to formulate hypotheses. It is clear that the formulation of hypotheses does not occur automatically but is usually preceded by many false starts, evaluational propositions, vague statements, etc.
3. As an aid in understanding hypotheses, we noted that they may be developed at different levels of concreteness, from fairly common-sense statements to the relationships between complex, abstract variables.
4. Making hypotheses is a creative act, but we can study such acts. We saw that hypotheses come from many sources, from the general emphases of our culture to the most individual of experiences.
5. Finally, we sketched a few criteria for *selecting* the more useful hypotheses and offered a few suggestions for improving those hypotheses which seem to be weak.

Such an outline at least offers the student a set of preliminary but useful notions for thinking fruitfully about research problems. Many studies fail at precisely this point, the development of a good hypothesis. On the other hand, the history of science gives innumerable examples to prove that great strides were made when someone asked the right question.

CHAPTER ONE

CLASS NOTES

I

The teachers and students who will use the materials contained in this selection of readings will undoubtedly possess varied intellectual interests and training. Consequently, we do not think there is one clearly preferred way to use these readings in the classroom, one "correct" approach which will insure both enthusiasm and understanding. Nonetheless, the following problems and illustrations may aid law students and their teachers in coming to grips with some major methodological issues in the social sciences. The emphasis here is not so much on reading or discussion as on performance; and to this end we have included a number of "exercises" at the end of each chapter which may be used as they stand or modified to fit the expertise and interests of the class.

II

At the beginning of these readings, we said that there is no proven method to teach people how to develop hypotheses. Nonetheless, we think it is possible to set up a series of tasks which have a good chance of helping the student to uncover ideas worth testing.

A. The first step is to have the student decide on the subject (or variable) that interests him, that he would like to understand in terms of what makes it vary. As an illustration, let us say it is the matter of academic performance in law school. We can call this the dependent variable--that is, the variable that presumably depends on some other variable or set of variables we wish to discover. The latter are usually referred to as independent variables. The task, then, is to discover what independent variables are related to--and possibly cause--variation in the dependent variable of academic performance in law school.

The first and simplest way to look for an answer is to use common sense and our own experience and intuition. We might guess that there is an intellectual aptitude or ability for the study of law that plays a large part in determining just how well a law student does in his course work. We might also guess that past academic performance also plays a part--perhaps reflecting an ability that partly overlaps a special aptitude for the study of law but is not conterminous with it. And, finally, we might suppose that academic achievement in law school transcends intellectual skills--there is a motivational factor, a drive to do well, that students for one reason or another possess to a greater or lesser degree.

We can now set these variables up in a matrix, as in Figure 1, and systematically examine the relationships among them. Each crossmark in a cell stands for a relationship between the variables taken two at a time.

The next step is to ask what relationships, if any, exist between these variables and *to state explicitly why we think the relationships exist in the way we have asserted*.[1]

---

[1] Since academic achievement in law school is our dependent variable, it might be thought that there is no benefit to be gained in trying to link the independent variables, such as motivation and aptitude for the study of law. But such a task helps to better understand why the independent variables have the influence that they have and how their interaction effects the dependent variable.

FIGURE I

	Academic Achievement in Law School	Aptitude for Study of Law	Past Academic Performance	Motivation
Academic Achievement in Law School				
Aptitude for Study of Law	X			
Past Academic Performance	X	X		
Motivation	X	X	X	

There is one more point to be considered in this part of this first exercise. Before we can test our assertions, to see if they correspond to empirical reality, we must *operationalize* our terms. The concept of operationalism has a long and complicated history; and it raises some very difficult issues in the philosophy of science.[2] Here, however, all we wish to make plain is that although defining every concept or variable we discuss would be a waste of time, there are key concepts which we must define as precisely as we can in terms of the aspects of the observable world to which they refer--and if such definitions are to be useful in research and are to be comparable from one piece of research to another, we will have to specify the operations by which we make our observations. (For example, what do we mean by the term "socio-economic status?" What empirical reality does it refer to? How can we measure that empirical reality?)

In the problem we are now dealing with, the academic achievement of the law student, we might define our terms as follows:

☐ Academic achievements will refer to the cumulative grade point average of the law student as recorded by that great keeper of records, the Dean's office.

---

[2] See, for example, Herbert Feigel, "Operationalism and Scientific Method," *The Psychological Review*, Vol. 25, No. 5, September, 1945, pp. 250-259.

☐ Aptitude for the study of law will refer to the student's score on the morning session of the Law School Aptitude Test, as administered by Educational Testing Service.

☐ Past academic performance will refer to the student's cumulative grade point average as an undergraduate.

☐ Motivation will refer to--what the reader of this book can create as a measure of the individual law student's drive to do well in law school. What behavior, what verbal responses, could serve as a good indicator of different levels of ambition for academic achievement?

It is possible, of course, to argue that grades are not a very good indicator of academic achievement, either in law school or in college, and that the LSAT is not a very good indicator of aptitude for the law.[3] We are free to construct another operational definition, if we wish--remembering, however, to pay attention to common sense, common usage, and the definitions used by social scientists in other research projects.

---

[3] At the end of these class notes, we have included two readings which bear on this issue. The first is an excerpt from the LSAT Handbook; the second is an article, "Law School Admissions: Science, Art, or Hunch?" by Robert R. Ramsey, Jr. which appeared in *The Journal of Legal Education*, 1959-60, Volume 12, pp. 503-520.

B. A second device for generating hypotheses involves a critical reading of the work of others who have studied the problem being examined. It is here we can find propositions that have been tested once but need to be tested again; explanations for observed results which sound plausible but have been subjected to empirical examination; inconsistencies or discrepancies which are puzzling and need further exploration; insights which are hardly more than vague hints and which need to be worked out and hooked to clear, specific empirical referents. Scientific understanding, after all, is to a large extent a matter of slow accumulation of tested and refined ideas. The brilliant, original hypothesis is a fine thing, of course--and it is essential for the real growth of a discipline. But the testing and development of existing ideas is also essential and fully deserves the effort it requires.

The two readings we mentioned before--the excerpt from the LSAT Handbook and the article by Robert R. Ramsey, Jr., will be useful here. Both contain numerous assertions which can be developed into hypotheses.

C. A third device involves looking at work that has been done in analogous areas. The use of analogies, in fact, is one of the most potent means for the discovery of new ideas, new insights. The situation which serves as an analogy may be very similar to the topic being examined or it may be quite different in a great many respects; in either event, the attempt to find similarities and differences between two phenomena has proven to be an invaluable aid in scientific discovery.

What, then, is similar to achievement in law school? One obvious and very similar situation is achievement in college. (We suggested earlier that achievement in college might reflect an ability which would influence performance in law school. Now we are using achievement in college in another way--as an analogous situation which may suggest other factors influencing performance in law school.)

There have, of course, been a great many studies of undergraduate performance. We will cite only one--_They Went To College_ by Ernest Havemann and Patricia West.[4] These authors found that two variables were related to performance in college--namely, political opinions and extent of extra-curricular activities. Students who called themselves _Independents_ tended to get higher grades than students who call themselves _Democrats_ or _Republicans_. (The study was made in 1947.) In addition, the authors found that students who engaged in many extra-curricular activities tended to get higher grades than those who did not. The question, then, is whether these factors influencing _undergraduate_ performance might also influence law school performance--and if so, precisely why.

---

[4] Ernest Havemann, Patricia Salter West, _They Went To College, The College Graduate in America Today_ (New York: Harcourt, Brace and Company, 1952)

We said this is a rather obviously analogous situation. In fact, there are a great many other situations which have been examined with regard to the determinants of achievement and which might help to generate ideas or hypotheses about achievement in law school. The factors influencing performance in the business world, in the Army, in voluntary associations--all of these and more may open up additional hypotheses.

LSAT Handbook
Educational Testing Council
1964

What the Law School Admission Test Measures

The development of the Law School Admission Test illustrates a number of these generalizations. One of its main functions is to assist in the selection of law students by providing information about probable future performance in law school. Obviously it cannot test law school applicants on knowledge and skills which will be acquired during their law school careers. It must, therefore, depend on indirect methods--on the measurement of general abilities underlying those of ultimate interest and on the measurement of related skills in contexts which are free of specialized knowledge.

A thorough analysis of what it takes to do well in law school--as defined by performance on first-year examinations--would result in a long list of abilities and personal qualities among which might be the following:

facility in acquiring information,
skill in logical analysis,
writing ability,
persistence,
originality,
effective study habits,
self-discipline,
ability to work well in a competitive atmosphere,
motivation to succeed.

The list could be extended indefinitely, but even these few examples will serve to illustrate some of the problems facing the designers of a test like the Law School Admission Test. Some of these qualities elude the precise operational definitions which are a necessary basis for successful test development. Among these are originality, self-discipline, ability to work under competition, and motivation to succeed. Undoubtedly these could be further analyzed, but when one asks exactly how a person possessing these qualities would behave in a concrete situation, difficulties arise. What seems like originality to one person is not so classed by another; self-discipline, while generally desirable, may be foreign to a student who excels by virtue of his original insights; conditions of competition similar to those existing in a law school are difficult to reproduce in a testing situation; and motivation to succeed is so vague a quality as it stands, and so complex a quality upon further analysis, that it becomes a subject for basic research rather than a quality which can be tested by existing methods.

It is true, of course, that through the operation of serendipidity all of these qualities and others may be indirectly measured to some degree in a test designed primarily to test other abilities. The Law School Admission Test is at present a full day's battery of tests during which the candidate performs numerous taxing tasks under high stress, and the mere fact that he appears for the test and completes it may imply something about his persistence, his self-discipline, his ability to work under competition, his motivation to succeed, and many other qualities not directly approachable.

Other qualities, such as effectiveness of study habits, could probably be given satisfactory operational definitions but raise serious problems of testing method. Some information could no doubt be obtained about them, but the method employed would probably be self-report which, under conditions of competitive examination, is notably undependable. Again, while such qualities are not directly measurable, information about them may be implied in the results of other types of measurement. One might assume, for example, that in most instances the student who has good undergraduate grades from a school that demands much of its students has probably developed reasonably good study habits, although this may be an erroneous assumption in the case of a student who is bright enough to achieve good grades with little effort.

It is apparent, then, that by no means all of the abilities and qualities which result in a successful law school career are susceptible of measurement. Some of them defy the precise definition in concrete behavioral terms which is a necessary preliminary step in test development. Others present problems of testing method, given the conditions under which tests must be administered to large numbers of candidates. There remains from the suggested list of abilities and personal qualities a small group for which reasonably exact definitions can be provided, for which useful thought imperfect test methods do exist, and which are of great importance to success both in the study and in the practice of law. These are skill in acquiring information, logical analysis, and writing.

The measurement of these abilities is diffused throughout the test. No one section of the test is exclusively related to any one of them. However, for convenience in presentation, it seems best to discuss the measurement of each of these abilities in connection with the specific sections of the test where it is most clearly displayed.

Robert R. Ramsey, Jr., "Journal of Legal Education,"
1959-60, Volume 12, pp 503-520.

LAW SCHOOL ADMISSIONS: SCIENCE, ART, OR HUNCH? *

Robert R. Ramsey, Jr. †

LAW school admissions committees have long been interested in more precise methods for estimating a student's chances for success in the LL.B. program, but only in the last decade has the growing admissions problem generated attention on a broad scale to the need and likely efficiency of rigorous and systematic evaluation procedures. The advent of selective admissions has, in particular, meant a new emphasis on prediction, for all selective methods for admission are essentially predictive. The committee which grants admission to one and denies it to another is, in effect, saying that it feels the former is better law school material than the latter, and its members must likewise feel that of those accepted, some will do better than others.

The problem of selective admissions was not peculiar to the late forties; it is a continuing one, and promises to be even more crucial in the years ahead. One can hardly overlook the great population expansion the nation has been experiencing in the last few years. The baby boom of the past two decades has already added millions to the class lists of the 'sixties and 'seventies. Nor is the problem solely one of quantity. An additional factor that renders the admissions situation even more acute is the noticeable improvement in quality of the applicants who each year seek entrance to our colleges and professional schools. Preliminary screening by counselors and advisers has aided greatly in reducing the number of marginal applications to be considered, but this very act has also made the selection task more difficult. Institutionally and individually, we can no longer afford Wigmore's trial-and-error method that "the way to find out whether a boy has the makings of a competent lawyer is to see what he can do in a first year of law studies." [1] Of those who now complete application, a very large percentage must be considered qualified in so far as experience in the past with individuals of similar qualifications has shown that they are likely to attain a satisfactory level of academic success.

---

* The author is indebted to Professor Paul S. Burnham, Director, Office of Educational Research, Yale University, and to Mr. Louis A. Toepfer, Vice Dean, Harvard University Law School, for helpful criticisms and suggestions concerning this article.

† Assistant Director, Office of Educational Research, Yale University.

[1] Wigmore, *Juristic Psychopoyemetrology—or How to Find Out Whether a Boy Has the Makings of a Lawyer*, 24 Ill.L.Rev. 454, 463 (1929).

Law schools have felt these pressures, and will feel them to an even greater degree, for not only are they fewer in number, but their facilities for handling expanded enrollments are generally more limited. Certain other considerations also give selection for professional training added importance: the cost to the individual and to society of a professional education, highly desirable personality qualities and abilities, probable ethics of practice, and possibility of social contribution.[2]

In view of these factors, it seems likely that explicit evaluation procedures will assume an increasingly important role in educational practice. Used wisely, they can contribute significantly to a realization of Kandel's plea of twenty years ago, even timelier for today's situation: more than ever before, the individual student must be protected as far as possible against the disappointment, disillusionment, and waste that arise from his being admitted to a college or professional school for which he is unsuited or inadequately prepared.[3] The professional school must also minimize the waste of its own limited resources by discriminating more finely between those who show promise and those who do not. The ultimate goal of any such program is the same at all levels: the right education for the right individual, a greater awareness of individual differences among students and of the individual characteristics of a particular student which make desirable for him a certain educational program as opposed to the several available to him. Hopefully, rigorous and systematic assessments of admissions procedures will enable both student and institution to capitalize more fully on the educational opportunities at their disposal. The present article reviews the law schools' experience of the past and present, and offers some suggestions for the future.

I

THE NATURE OF PREDICTION PROCEDURES

The necessary components of any prediction procedure are a criterion and at least one predictor. The most fundamental, and perhaps the most difficult, problem in any prediction study is the choice of some standard of subsequent performance against which to validate the procedure in question, for it is axiomatic that any such research is only as good as the criterion used. In selecting students for law school, two kinds of performance suggest themselves as possibilities for the criterion to be predicted: success in legal practice and success in law school. The former represents what Thorndike calls the "ultimate" criterion, the complete final goal of professional education.[4] But as he himself is quick to remind us, such a criterion is rarely, if ever, available for use in psychological research. This is certainly the case with respect to the professions, for the securing of a satisfactory on-the-job criterion has, to date, been practically impossible. Because the duties and responsibilities of the professional person vary from time to time and from place to place, success in professional work has proved too elusive and relative a concept to be of use in the validation of more precise prediction techniques. Such criteria of long-term proficiency as length of service, income derived from work, and rank or professional standing are helpful, but they too have their limitations, in that it is not entirely accurate to equate success with any one or particular pattern of these.

---

[2] English & Nagy, *Professional School Admissions*, in 2 HANDBOOK OF APPLIED PSYCHOLOGY 453 (1950).

[3] I. L. KANDEL, PROFESSIONAL APTITUDE TESTS IN MEDICINE, LAW AND ENGINEERING v–vi (1940).

As a result, the researcher is almost always thrown back upon substitute, working criteria of an "immediate" or "intermediate" nature, which he judges, either in terms of rational analysis or in terms of empirical evidence, to be related to the ultimate criterion with which he is most fundamentally concerned. Prediction studies with law students have historically relied on "immediate" measures of success, generally first-year grade average. It is important to remember that all so-called "immediate" or "intermediate" criteria remain partial, since at best they give only an indication or an approximation of the ultimate goal towards which the selection is directed.

An acceptable working criterion should possess the following qualities: (1) It must be relevant to the ultimate goal, and this relevance should be based on actually determined correlations; (2) it must have some reliability, including the power to discriminate between individuals; (3) it must be as objective as possible; and (4) it must be practical for purposes of convenience, economy, and comparability.[5] Anyone who has had occasion to be concerned operationally with the choice and use of criterion measures realizes that few of those currently available satisfy all of these requirements. Graduation, over-all grade average, first-year and first-semester grade averages, and achievement in particular courses, though imperfect, are useful as long as the investigator recognizes their limitations and makes allowance for the same in his final interpretations.

Most law school prediction studies have used course grades as both predictor and criterion, correlating undergraduate college grades against law school grades. The continued widespread use of grades in prediction studies is due in large part to the fact that they are conveniently available and *appear* to possess a rather satisfactory degree of reliability. However, it is widely recognized in the psychological and statistical world that "grades do not represent the universal and unvarying values which their numerical or literal definiteness often implies."[6] The unreliability of instructors' grades, based, as many contend, on "some esoteric combination of improvised testing, dazzling intuitions, and the persistence of the student attending classes,"[7] has long been a controversial topic and a continuing problem. Professor Grant has recently proposed in this *Journal* one method for the establishment of fixed, discernible standards in law school grading, in the hope of reducing "the incredible variance now caused by our present personal and incompatible grading standards that so distorts the significance of our grades to the injury of our schools and our students."[8] College and university grading methods have undergone favorable modification in recent years, but there still remains too much subjectivity in many grading situations. Some feel that this is due to a lack of any defined common denominator upon which to base grades, and they emphasize that it will continue "as long as each teacher has a set of standards, individually defined and reflecting a somewhat unique set of objectives."[9] Unless a criterion is free of such arbitrary elements, satisfactory predictions will be difficult, if not impossible, to make.

---

[4] R. L. Thorndike, Personnel Selection 120-24 (1949).

[5] Rulon, *The Criterion*, in E. F. Lindquist (Ed.), Educational Measurement 634 (1951).

Most prediction studies have used the average or summary grade, and there is the feeling in some circles that such an over-all evaluation of the student has limited validity for prediction purposes. The chief objection is that such a measure fails to indicate the degree to which each important goal of instruction has been achieved by the student. Notable among these critics is Furst, who comments that,[10]

> Such evaluations have some usefulness in prediction studies but, in general, suffer from the limitation of not being analytic . . . . The weakness of an average is that it is always somewhat artificial; it implies uniformity where variability is the rule. Instead of describing the pattern of achievement over the various instructional objectives, it yields only a conglomerate the parts of which are rather nondescript.

While no one can deny that grades or a summary of them represent varying degrees of achievement in many different kinds of courses taught by various kinds of instructors with different standards of measurement (in the process, giving rise to a certain amount of unreliability and bias), the general feeling is that sufficient reliability and validity remain in the final composite index to make them useful for prediction purposes. Their continuing value stems from the the fact that they represent a combination of ability and motivational factors operating in much the same way as they probably will in professional school. The advantages of these circumstances seem to outweigh the factors that tend to reduce their reliability and validity otherwise.

---

[6] J. G. DARLEY ET AL., THE USE OF TESTS IN COLLEGE 13 (1947).

[7] Sarbin & Bordin, *New Criteria for Old*, 1 EDUC. & PSYCH. MEASUREMENT 173, 180–81 (1941).

[8] Grant, *Justice in Grading*, 9 J.LEGAL ED. 212 (1956).

[9] Sarbin & Bordin, *supra* note 7, at 174.

[10] Furst, *Theoretical Problems in the Selection of Students for Professional Schools*, 10 EDUC. & PSYCH. MEASUREMENT 637, 648 (1950).

In recent years, standardized tests have gained increasing acceptance as useful predictors, so that today nearly all professional schools require applicants for admission to sit for specially-constructed aptitude tests, supervised and administered by the respective professional associations or independent testing organizations. General achievement and personality inventories have also been used from time to time, but the correlations between scores on these scales and course grades in professional schools have been consistently low and not of sufficient magnitude to be useful in individual prediction when selecting applicants for admission.[11]

In order to interpret the results of prediction studies an index of the relationship of predictors to criterion is necessary, since some factors are presumably better indicators of a student's probabilities of success in a prescribed program of professional studies than others. The usual method of indicating this relationship has been the coefficient of correlation. The larger the correlation coefficient (the possible range is – 1.00 to + 1.00),[12] the more we assume the predictors and criterion to have in common, high performance on the former reflecting substantially the same factors that apparently account for academic success in professional school.

In the final analysis, the use of any prediction technique in an admissions operation will depend upon how well it serves the objective of the particular professional school. Each school must determine its own admissions policy. If it proposes to admit a large number of students with the expectation that a considerable number will fail, then the predictor need not be so predictive. On the other hand, if the aim of the school is to pick out only those successful students at the start (as is usually the case in these days of many applicants and limited facilities), then the most valid predictor available is still inadequate.

---

[11] See Thompson, *Personality and Interest Factors in Dental School Success*, 4 EDUC. & PSYCH. MEASUREMENT 299 (1944); Melville & Frederiksen, *Achievement of Freshman Engineering Students and the Strong Vocational Interest Blank*, 36 J. APPL.PSYCH. 169 (1952); Glaser, *Predicting Achievement in Medical School*, 35 *id.* at 272 (1951).

[12] Obtained correlation coefficients rarely exceed + 0.75, and the following is a rough guide to their evaluation *in terms of experience to date*:

0.00 – 0.20	Very Low
0.20 – 0.40	Low
0.40 – 0.50	Average
0.50 – 0.70	High
0.70 – 1.00	Very High

II

REVIEW OF LAW SCHOOL EXPERIENCE

As in most other professional fields, an increasing number of applicants and the persistence of a high percentage of failures led law schools to seek more selective methods of admission. Thirty years ago, there was practically no selective admissions policy, and during this trial-and-error era, almost any college graduate could enter the law school of his choice, provided he was of good character and could pay the bills. As a result, the first-year failure rate at many schools was quite severe; and at one, a professor is reputed to have suggested to members of an entering class that they "look at the man to the right of you, look at the man to the left of you; one of you three won't be here next year." [13]

The approach of the earliest law school studies was to try to determine the degree of relationship between the quality of the student's prelegal work and his achievement in the law curriculum.[14] In other words, to predict marks, use marks. Studies of this type continue to the present, and the findings to date have shown that the undergraduate academic record is a fairly reliable predictor of subsequent achievement in law school.[15] Many legal educators maintain that the quality of the student's prelegal college work is the best single indicator of law school success, and they are supported in this conviction by a sizable body of research evidence. In many cases, the college average is not so valuable in predicting the actual achievement in law school as it is in forecasting potential failures.

It is interesting to note that in some instances, accuracy of prediction tends to increase if it is based on prelegal grades received at the same institution where the prospective student intends to pursue legal training. It also appears that frequently the quality of a student's academic record during the period immediately prior to entrance into law school is of greater predictive utility than that of the first years of preparation. Studies at Yale Law School have found a rather high correlation between undergraduate junior year average and law school performance, and the suggested explanation is that this average "probably represents achievement at a level of motivation and work-adjustment more comparable to that in the study of law than is the case with the two preceding years of undergraduate preparation."[16]

---

[13] STUDYING LAW AT THE HARVARD LAW SCHOOL 22 (1955).

[14] D. B. STUIT ET AL., PREDICTING SUCCESS IN PROFESSIONAL SCHOOLS 48- (1949).

[15] A mimeographed tabular summary of published correlation studies between prelaw grade averages and law school achievement up to 1948 is available upon request to the author (Box 2177, Yale Station, New Haven, Conn.).

As one means of combatting the high failure rate, the minimum requirements for admission were raised from two years of college to college graduation. But, as Kandel pointed out,[17]

> this provided no assurance of success in legal studies. Mortality continued to be high and could not be accounted for by the lack of general intelligence; students with good college records were also found among the failures in law schools.

In the middle 'twenties, it was concluded that success in law school required specific kinds of intellectual capacities, and a study was undertaken to ascertain the capacities a lawyer supposedly requires for his work. On this basis, early law aptitude examinations were designed to measure such things as power of analysis, recognition of the crucial point in a mass of facts, ability to memorize quickly, and ability to assimilate what has been learned. Alertness, verbal facility, and the possession of a large vocabulary also appeared to be important in these early studies.

Prior to 1948, at least seven specially constructed legal aptitude tests were used by various institutions as one means of selective admission to their law schools. With the exception of the Ferson-Stoddard examination, however, use of these tests was generally restricted to the institutions at which they were developed.[18]

The Ferson-Stoddard Law Aptitude Examination, constructed by a law school dean and a psychologist, was based upon the following abilities, as representing, in the aggregate, a certain degree of law aptitude: (1) capacity for accurate recall, (2) reading comprehension, (3) reasoning by analogy, (4) reasoning by analysis, and (5) skill in pure logic. Most of the early studies showed that every part of the test was positively correlated with first-year and first-semester grades in law,[19] but subsequent experience revealed that it probably did not discover the high-grade students as accurately as the low-grade ones.[20]

The Ferson-Stoddard Examination was opposed from its very inception by some legal educators, and among its most vociferous opponents was Professor Wigmore of Northwestern. Conducting his own study with fifty student volunteers, he concluded that the test[21]

---

[16] Crawford & Gorham, *The Yale Legal Aptitude Test*, 40 YALE L.J. 1237, 1244 n. 8 (1940).

[17] KANDEL, *op. cit. supra* note 3, at 34.

[18] A mimeographed tabular summary of published correlation studies between law aptitude test scores and law school achievement up to 1948 is available upon request to the author.

[19] Stoddard, *Ferson and Stoddard Law Aptitude Examination: Preliminary Report*, 6 AM.L.SCH.REV. 78 (1927).

[20] Witham, *Legal Aptitude Tests*, 25 ILL.L.REV. 448 (1930).

> as a basis for juristic psychopoyemetrological prognostication of individual aptitude . . . had no substantial practical value for the purpose intended, viz., predictability as to ultimate high capacity of a particular individual for law studies.

Not having utilized accepted statistical techniques such as correlation, median, and deviation, he did concede, however, that some "psychopoyemetromethodological" error might be discovered in his methods; and unbiased evaluation of all available validity data on the Examination suggests that it did warrant more credit than his conclusions admitted.

Columbia Law School experimented with law aptitude tests as early as 1921. With the assistance of Professor E. L. Thorndike, a law "capacity" test was prepared "to test the student's ability to work effectively with abstractions and symbols, a capacity necessary to do the kind of work required of law students."[22] Although it was not designed to determine general mental ability, its sponsors claimed that it was so constructed that it was next to impossible for a student not fitted for the study of law to obtain a high score. The test was required of every student admitted to the first year from 1928 to 1936, but in the latter year, the admissions committee was given the power of discretion to admit to the first-year class without the requirement of the capacity test.

An account of Columbia's experience with the test may be found in an appendix to the Dean's report for 1937.[23] Although the capacity test was not a certain index of law school achievement (*i. e.,* able students occasionally did make a low score), the findings with 1980 students during the nine-year period (coefficient of correlation, + 0.38) did justify its use as a predictor of law school performance. A slightly higher correlation (+ 0.41) was obtained between college average and first-year law grades. In the operation of its selective admissions process, Columbia personnel found that the college average and the capacity test score acted as valuable checks upon each other, particularly when the applicant had a low placement in either. Thus, it was discovered fairly early that law aptitude test results may actually improve predictions based on college grades alone only slightly, but they can serve the valuable additional function of providing a common base for evaluating college records. Such verification can be especially helpful to an admissions committee that

[21] Wigmore, *supra* note 1, at 463.

[22] REPORT OF THE DEAN OF THE LAW SCHOOL FOR THE PERIOD ENDING JUNE 30, 1937, 20 (1937).

[23] Jacobs, *Report of the Faculty Committee on the Selection of Students*, appendix to the REPORT OF THE DEAN OF THE LAW SCHOOL FOR THE PERIOD ENDING JUNE 30, 1937, 19-58 (1937).

must consider applicants from a large number of relatively "unknown" colleges.

Columbia and Yale have used legal aptitude tests continuously over a longer period than any other law schools. Yale first experimented with the Thorndike CAVD (Completion, Arithmetic, Vocabulary, Directions) Test, and a much shorter test of twenty minutes constructed by its own Professor Chapman. But these were not specific tests of legal aptitude, and after several years of unsatisfactory experimentation, the assistance of Crawford was enlisted to design a new instrument, the Yale Legal Aptitude Test. This was a test of ninety minutes and was composed of verbal comprehension questions, logical inference and analogy problems, and legalistic material.[24] Although legal problems and terminology were involved, knowledge of law was not required by any part of the test; all the professional information needed to answer the questions was provided in the directions of the test. In short, the Yale test was planned "to reveal whether or not the student possesses a capable though legally naive mind." [25]

Validity studies involving the Yale Legal Aptitude Test revealed higher correlations between test score and first-year law grades (+ 0.55) than corresponding measures reported for the Ferson-Stoddard (+ 0.46) and Columbia (+ 0.38) law tests, and Yale authorities found that it was particularly helpful as a common denominator in checking on the validity of undergraduate records.[26] The experience of the University of Wisconsin with the Yale test, however, indicated that it added little to prediction of grades earned in its law school; and it did not appear to be of quite as much service as the straight aptitude test,[27] supporting Crawford and Gorham's conclusion that [28]

> any prognostic instrument which, like the Legal Aptitude Test, has been validated and successively revised with reference to a particular curriculum and population is likely to find its usefulness limited to situations closely analogous to those which have governed its own development.

Thus, in adopting existing tests for new situations, one should be prepared for a certain predictive shrinkage as the by-product of a ready-made versus a tailored solution.

The University of Minnesota tried a number of different tests during the 'thirties, but unlike Yale, was definitely not satisfied with its attempts to measure legal aptitude.[29] Using original and borrowed tests, it found that the most successful single instrument was the University of Minnesota Relevancy Test in Law, a test of the relevancy of propositions of law to stated questions of fact. At various times, its legal aptitude battery was also composed of the Wesley Social Terms Test, Legal Generalization Test, and Prosser Reading Tests. After five years of experimentation (1933–37), during which time the predictive validity of the tests varied incomprehensibly from year to year, the use of aptitude tests was discontinued by the Minnesota Law School. Experience had shown that by far the best predictor for its own situation was a prelaw college average adjusted to take account of the fact that some schools are weak and inferior.[30]

---

[24] Crawford, *The Legal Aptitude Test Experiment at Yale*, 7 AM.L.SCH.REV. 530 (1932).

[25] Crawford & Gorham, *supra* note 16, at 1240.

[26] Gulliver, *The Use of a Legal Aptitude Test in the Selection of Law School Students*, 9 AM.L.SCH.REV. 560 (1940).

[27] Husband, *Prediction of Law School Success*, 2 WIS.L.REV. 285 (1939).

[28] Crawford & Gorham, *supra* note 16, at 1248.

The law schools of the Universities of California and Michigan also experimented with the predictive value of legal aptitude tests at one time or another, but very little has ever been published regarding their findings. The California Legal Aptitude Tests were prepared locally, and correlation coefficients of + 0.52, + 0.64, and + 0.61 between test scores and law school grades were reported for three years.[31] Michigan reported a correlation of + 0.60 between scores on its own legal aptitude test and first-year law grades for 1940.[32]

In 1941, an experimental edition of the Iowa Legal Aptitude Test appeared as "a new test of scholastic aptitude for legal studies which would incorporate the best features of existing tests of this type, and which could be made available for general use by schools of law and college personnel counselors." [33] Subsequent validity data led to a 1946 revision composed of the following seven parts: analogies, reasoning, opposites, relevancy, mixed relations, memory, and common legal information (which tested a candidate's knowledge of subject matter to a limited extent).[34] Although studies show that the Iowa Test was probably not a great deal better, on the average, than the undergraduate grade record for predicting academic performance in law school, its primary value derived from its availability for use in the counseling of individuals whose undergraduate preparation was of marginal quality.

---

[29] Prosser, *Comments on Dean Gulliver's Paper from Proceedings of the 37th Annual Meeting of the AALS*, 9 AM.L.SCH.REV. 658 (1940).

[30] Douglass, Luker, & Lovegren, *Prediction of Success in the Law School*, in UNIVERSITY OF MINNESOTA STUDIES IN PREDICTING SCHOLASTIC ACHIEVEMENT 46 (1942).

[31] Kidd, *Comments on Dean Gulliver's Paper from Proceedings of the 37th Annual Meeting of the AALS*, 9 AM.L.SCH.REV. 663 (1940).

[32] Communication from J. E. Tracey to W. M. Adams, as reported in Adams, *Prediction of Scholastic Success in Colleges of Law: I. Experimental Edition of the Iowa Legal Aptitude Test*, 3 EDUC. & PSYCH. MEASUREMENT 291 (1943).

[33] *Id.* at 296.

[34] Adams & Stuit, *The Predictive Efficiency of the 1946 Revision of the Iowa Legal Aptitude Test*, 9 EDUC. & PSYCH. MEASUREMENT 23 (1949).

Not satisfied with the predictive efficacy of existing instruments, a group of law schools approached the College Entrance Examination Board in 1947 regarding the possibility of developing a legal aptitude test which would be available nationally several times a year. (As early as 1939, at the 37th Annual Meeting of the AALS, Kidd had proposed uniform national aptitude tests for the legal profession, in the hope that better tests devised by experts might reduce materially the enormous wastage then existent.[35]) The result of these negotiations with the CEEB was the Law School Admission Test (LSAT), and an LSAT Policy Committee, composed of legal educators and admissions officers from representative law schools, was established from the start to advise the technical experts in carrying on a continuous program of test development, pretesting, item-type validation, and test validation.

From the very beginning, the Policy Committee has emphasized that the LSAT "should be an aptitude (ability) test as free as possible of items requiring specific knowledge or detailed information, yet predictive of scholastic achievement in law school." [36] The intention of the test, now required as part of the admissions procedure at seventy-six law schools, is not to measure a student's prelaw scholastic achievement, nor his motivation or willingness to apply himself. Nor does it purport to predict success in legal practice. Its ultimate function is to "predict only what an examinee can do in the study of law if he uses his ability well—what his performance is likely to be if he applies himself." [37] Questions on the test are, therefore, designed to measure the individual's capacity to read, to understand, and to reason logically with a variety of verbal, quantitative, and symbolic materials.

The results of validity studies with the LSAT vary from school to school. In some cases, prelaw record is the better predictor; in many instances, the LSAT is superior as a single predictor. Studies conducted by the Educational Testing Service on 1948 entrance at fourteen law schools showed a range of coefficients of correlation between LSAT score and first-year law grades from + 0.18 to + 0.65, with the mean falling at + 0.41. The corresponding mean coefficient for the undergraduate grade average was + 0.37. Studies carried out by ETS subsequent to 1948 tend to confirm these results.[38]

Most law schools have found that the best prediction technique currently available is some combination of prelaw record and LSAT score, for such a combination has proven superior in almost all instances to either predictor used alone.[39] Such combinations, based on statistically determined "best" weights of the predictors, have yielded multiple correlation coefficients as high as + 0.60, a significant improvement in predictive efficacy over that of the predictors employed separately.

---

[35] Kidd, *supra* note 31, at 664.

[36] Johnson, *Validation of Professional Aptitude Batteries: Tests for Law*, 1950 ETS Invitational Conference on Testing Problems, Proceedings, 31 (1951).

[37] A. P. Johnson, Marjorie A. Olsen & J. A. Winterbottom, The Law School Admission Test and Suggestions for its Use 9 (1955).

[38] Communication to the author from J. A. Winterbottom (1959).

III

Implications for the Future

Errors in the prediction of law school performance are still fairly high, and it is apparent to all that no prediction technique has as yet even approached infallibility. Correlations in the 0.50's and 0.60's are useful, but they are low enough to impress upon us that our attempts to predict law school achievement are still more in the realm of hunch than of science. Looking to the future, let us consider some of the recurring pitfalls of prediction, as well as the prospects for their resolution.

Correlations between first-year law grades and earlier legal aptitude test scores were, in many instances, larger than the more recent ones involving the LSAT. Does this mean that accuracy of prediction is, in fact, inversely proportional to the technical perfection of our measuring instruments? On the contrary, paradoxical as it may seem, Williamson is essentially correct when he suggests that we should "expect further lower correlations, not higher ones, as indications of efficiency and effectiveness."[40] We are here confronted with the peculiar statistical phenomenon that increased reliance on prediction renders correlation itself a less useful index. The elimination of low prediction cases through the admissions process increases the homogeneity of ability in the accepted group and thereby yields lower coefficients of correlation between predictors and criterion than would otherwise be the case in the more heterogeneous range of the total, unselected group. As more and more of the effectiveness of predictors is "used up" in the selection process, it becomes increasingly difficult to demonstrate statistically the real validity of the predictor. It is even possible that the correlations may be zero although the instrument is valid for selection purposes.[41]

Admissions officers might view any such apparent decline in the validity of their selection techniques as analogous to the situation of our currency the longer it remains in circulation. While its appearance may deteriorate noticeably during the period of its use, its inherent value remains unchanged; and in times of an expanding economy—just as in these days of growing numbers of applicants—its importance to the prosperity of the situation is greater than ever. It is part of the lawyer's orientation that the facts are not always apparent on the surface. The seeming definiteness of a statistic should not lead him to sacrifice his professional skepticism in other areas.

---

[39] For an example of the multiple approach as used at one law school, see Burnham & Crawford, *Law School Prediction at Mid-Century*, 10 J.Legal Ed. 189, 195-96 (1957).

[40] Williamson, *The Decreasing Accuracy of Scholastic Predictions*, 28 J.Educ. Psych. 1, 13 (1937).

[41] Travers & Wallace, *Inconsistency in the Predictive Value of a Battery of Tests*, 34 J.Appl.Psych. 237 (1950).

The continuing widespread use of course grades as both predictor and criterion has not ipso facto removed the obscurity that has so long characterized these academic devices. If they are to contribute to the prediction process as most of us hope and expect, their nature and origin must be subjected to closer scrutiny than has generally been the case in the past. Grant has outlined the problem with respect to law school grades,[42] and who among us has not paid at least lip service to purported differences in undergraduate departmental standards and the varying levels of ability represented by students who are attracted by them? Heightened accuracy in prediction is dependent upon improvement in institutional grading methods, but in the meantime, most law schools could profitably give more attention to differential techniques for evaluating applicants according to undergraduate college and major. If consistent differences in law school performance are observed among students from different undergraduate institutions, or from different colleges or departments within a single institution, then individual institutional corrections reflecting these differences in performance should be incorporated in future predictions.

Institutional "corrections" in the prediction procedure may be handled statistically in one of two ways, depending on the nature of the particular situation. The larger law schools, with a history of several students from certain "feeder" colleges in each entering class, may devise a separate prediction procedure for each undergraduate institution. In the author's opinion, a sample of at least thirty cases is desirable in order to effect any semblance of statistical reliability, and in nearly all law schools, large or small, such samples will generally have to be assembled from more than one entering class. Since conditions governing predictors may vary over a period of time, cases upon which the prediction procedure is built should be drawn from the minimum number of classes necessary to obtain acceptable sample size. Ordinarily samples of thirty to one hundred cases are satisfactory if they are drawn from no more than five successive classes.

---

[42] Grant, *supra* note 8.

Enrollments in the smaller law schools usually do not include sufficient numbers of students from any one undergraduate institution to make the derivation of separate prediction procedures statistically feasible. Records of students from several or all of the colleges represented may be combined to contribute to one or more group procedures. But even in this group approach, differential considerations for individual colleges can still be applied when warranted. How is this accomplished? At least one law school (and I suspect others) has relied on a "college grade adjusted" concept, consisting of three basic steps: (1) The undergraduate grade averages of all students from a particular college are transmuted to equivalent indices on the particular law school grading scale for which prediction is desired; (2) the means of the transmuted undergraduate and law school grade averages for these students are obtained; and (3) their difference is the correction to be applied to individual undergraduate grade records in making future predictions. For example, if the difference between the mean undergraduate grade average for all students from College *X* at Law School *Y* during the past five years and the mean of their first-year law grades during the same period is −7, then future applicants to Law School *Y* from College *X* should have their computed undergraduate grade averages increased by a factor of + 7 before prediction is accomplished.

A third major aspect of the law school admissions and prediction problem is the need for a program of continuing research and evaluation *at each school*. Studies supported by the LSAT Policy Committee have provided useful information on the total applicant group, as well as pilot studies at selected law schools. But ETS itself has repeatedly emphasized the limitations of these more general data for comparing any given individual with his competitors for admission to a particular law school.[43] The findings at one law school cannot be transferred unaltered to another institution without the possibility of considerable loss in validity. As shown in the Minnesota experience,[44] the best single predictor for one law school may be entirely negligible in its contribution to another. All of this reinforces Stuit's observation that "a test may have many validities, depending on the population and criterion of success with which it is used," [45] and points to a major deficiency in present law school admissions operations: the paucity of local validity studies.

It must be remembered that in practice, predictors never function in a statistical vacuum. Since external conditions and the law school's own educational philosophy have an acknowledged influence on their performance, each law school must determine the usefulness of predictors for its own purposes by establishing local norms and validities. The conclusions offered by any national or over-all study should always be regarded as "guideposts for local research rather than as generalizations valid for universal application."[46]

---

[43] Johnson, Olsen & Winterbottom, *op. cit. supra* note 37, at 64.

[44] Prosser, *supra* note 29.

[45] Stuit, *Significant Trends in Aptitude Testing*, in E. G. Williamson (Ed.), Trends in Student Personnel Work 76 (1949).

Once local validities have been established, they should be reviewed periodically.[47] The validation study of the relationship between prediction and performance is never finished as long as the particular procedure is being used as an integral part of the admissions operation. As in industry, there must be continuing quality control. Mere inference from previously obtained data is not enough, for just as the average ability level of the applicant group varies from school to school, it may also vary from year to year at the same school. In 1951, with the Korean conflict claiming a large percentage of the college-trained population, one law school found itself admitting a lower level of ability than has been sufficient to gain acceptance since that time. In fact, the ability level of its entering classes has subsequently risen to such an extent that many of those gaining admission nine years ago (and having since graduated) would not now survive the initial screening.

The findings reviewed here indicate that ability actually accounts for only a relatively small percentage of the variability of law school performance. It appears likely that as much as fifty per cent of the variance must be accounted for by other human attributes. Reviews of the literature on scholastic success[48] have repeatedly emphasized that ability, effort (drive or motivation), and circumstances (personal, social, economic, etc.) are *all* crucial factors in academic performance; yet, the most serious gap in our knowledge to date continues to be the assessment of these nonintellectual influences. The difficulty is probably both a function of their specifically nonintellectual nature and the fact that methods of obtaining such information are not so accurate as the more strictly objective measures of intelligence and achievement. The notion that "personality tests" are little more than a cheat and an intellectual fraud is still popular in some circles; and in many cases, this attitude is an almost insuperable hurdle to obtaining meaningful data in this area.

---

[46] Feder, *Student Personnel Work-I. Student Population*, in W. S. MONROE (ED.), ENCYCLOPEDIA OF EDUCATIONAL RESEARCH 1293 (1950).

[47] At Yale Law School there is a systematic validity study of the entire prediction procedure every three years, based on data from the five classes that have most recently completed the first-year curriculum. Burnham & Crawford, *supra* note 39. Law schools with larger student populations may properly decide to review all or portions of their procedure more frequently.

[48] Harris, *Factors Affecting College Grades: A Review of the Literature, 1930–1937*, 37 PSYCH.BULL. 125 (1940); Durflinger, *Prediction of College Success—A Summary of Recent Findings*, 19 J.AM.ASSOC.COLLEGIATE REGISTRARS 68 (1943); Borow, *Current Problems in the Prediction of College Performance*, 22 *id.* at 14 (1946); Travers, *Significant Research on the Prediction of Academic Success*, in WILMA T. DONAHUE, CLYDE H. COOMBS, & ROBERT M. W. TRAVERS (EDS.), THE MEASUREMENT OF STUDENT ADJUSTMENT AND ACHIEVEMENT 147 (1949).

Results from personality research supported by the LSAT Policy Committee have varied and generally have not shown any useful relation to the criterion of first-year law grades.[49] Nor have others succeeded in reducing significantly the unexplained variance in law school performance through psychometric evaluation of non-intellectual factors. A sociocultural approach was recently employed by the author [50] to study the performance of a sample of Harvard law students, and the results, while not immediately transmutable to a series of prediction equations, may, nevertheless, suggest new directions for research. Using the formulations of Kluckhohn [51] and McArthur [52] with dominant and variant value orientations, the analysis focussed on certain major institutional factors common to every individual's sociocultural matrix and which previous research had established as important bases, overtly or covertly, for differentiation in other areas of his behavior. The major hypothesis of the study was that individual differences in cultural background are, on the whole, reflected in differences in academic performance in college and law school.

Seven institutional variables (religious denomination, ancestral background, type of secondary school, father's occupation, family income, community, and region of residence) were investigated. Appropriate statistical procedures were utilized to ascertain if different groupings of these variables, singly and in combination, for two first-year classes at Harvard Law School showed cluster tendencies in a three-dimensional space defined by undergraduate college grade average, LSAT score, and first-year law school grade average.

Significant differences in academic performance were observed between groups within five of the institutional variables (religion, occupation, school, income, and region), and combinations of these single variables generated more highly significant differentiation in academic performance than any one of them alone. In the absence of significant and consistent LSAT mean differences, the observed grade differences were ascribed to the influence of nonability factors operating as products of dominant and variant value orientations (subcultures). In the most statistically significant of the combinations ("Religion+School+Occupation+Income"), certain cluster tendencies were evident, and it was demonstrated that Kluckhohn's "Being," "Being-in-Becoming," and "Doing" classification of personality types could be meaningfully imposed on the observed differences in academic performance. Differences among "School+Occupation+Income" and "School+Income" combinations also supported the utility of this trichotomous taxonomy of academic subcultures.

---

[49] See, *e.g.*, JOHN R. HILLS & WALTER J. RAINE, INTERRELATIONSHIPS AMONG SELECTED PERSONALITY, APTITUDE, AND ACHIEVEMENT MEASURES AT ONE LAW SCHOOL: A PRELIMINARY STUDY (1958).

[50] R. R. Ramsey, Jr., A Study of Cultural Influence on Academic Performance In College and Law School (unpublished thesis in Harvard University Library, 1959).

[51] Florence R. Kluckhohn, *Dominant and Variant Value Orientations*, in CLYDE KLUCKHOHN, HENRY A. MURRAY & DAVID M. SCHNEIDER (EDS.), PERSONALITY IN NATURE, SOCIETY, AND CULTURE 342 (1953).

[52] C. McArthur, *Personalities of Public and Private School Boys*, 24 HARV.EDUC. REV. 256 (1954); *Personality Differences Between Middle and Upper Classes*, 50 J. ABNORM.SOC.PSYCH. 247 (1955); *Upper-Class Intelligence as the Critical Case for a Theory of Middle-Class Bias*, 4 J.COUNSEL.PSYCH. 23 (1957).

Observed differences in academic performance were entirely in the direction of earned grades, with a notable discrepancy in favor of college grades. The complete lack of significant differences in LSAT performance was not too suprising when one considered that the LSAT, as a limited three-hour test of ability and aptitude, is much less susceptible to the sustained influence of nonintellectual factors than grades. But the absence of more significant differences in first-year law school performance was surprising.

The leveling of college grade differences in law school emphasizes again the complexities we face in bringing system and order to our law school selection procedures. The "Morning Glories" and "Late Bloomers" will probably always frustrate our best efforts of law school prediction, but we will have moved forward if we increase our awareness and understanding of the etiology of their occurrence. In the present study, one speculates that the improved performance of students from a "Being" background, typically characterized by a cultural and intellectual atmosphere in which the emphasis is on the flexible manipulation of ideas (or learning for its own sake), derives at least in part from their better training to deal with new or original situations than their "Doing" opposites, whose prior conditioning has focussed on the external symbols of success (grades) without commensurate attention to the more important underlying processes. The variability among cultural groups of the meaning of academic performance underscores a real danger in the application of any single, absolute standard or approach in the evaluation of academic behavior of students from different cultural backgrounds. The limited evidence of this study indicates that products of different cultural orientations perceive their academic roles differently, and further, that these very same perceptions may vary at different stages of their educational careers.

The need for more research to develop and evaluate different law school selection procedures has never been greater. Once validity studies within law schools have determined the most efficient use of intellectual and non-intellectual factors for the particular situation, an exchange of information relevant to the process (whether successful or unsuccessful) must follow if we are to bring order to the broader landscape. Only by pulling apart, *and* together, will we extricate ourselves from the quicksands of hunch and approach the foundations of science. In the meantime, we must live enviously with that rare and inexplicable creature, the admissions artist, whose persistent insights dazzle us all, himself included.

CHAPTER TWO

THE DESIGN OF RESEARCH

I

Designing a research project is best seen as a series of choices in which limited resources of time, money, and technical capability must be weighed against what we would like to do. It is almost always true that we cannot do exactly what we would like and that our plans must be compromised, our testing of hypotheses rendered less than perfect.

We will assume that the researcher has managed to hammer out the major questions he seeks to answer. Now he must decide what empirical data he needs to test his hypotheses, how he will collect those data, and how he will analyse those data once he has them. The fact that we must plan what we are going to do with our data even before we have them in hand is worth stressing, for it underlines the idea that research is not a neat sequence of distinct steps but is instead a set of interdependent parts which are often undertaken simultaneously. Thus, our plans for the analysis of data will frequently help us to decide what kind of data we need; and the kind of data we plan to collect will play a major plan in determining the kind of analysis that is possible. In any large-scale research project, in fact, no matter how well designed, we will find new hypotheses being developed near the end of the work, new kinds of data suddenly seen as essential and collected in haste, and so on. It is this view of research as a complicated set of interconnected events which is emphasized by Professor Selltiz, *et al.*, in the selection describing and illustrating the research process. They take pains to point out that the way research is reported in monographs and articles in professional journals is not the way it is carried out in reality, in a great many instances; and there is a great deal that is done that is very often not mentioned at all.

I I

Much of the work that is carried out today under the name of social science research is based upon interviews and questionnaires. Such a heavy reliance on one kind of data is, we think, a mistake; and we agree with the authors of *Unobtrusive Measures* when they say that, "We lament this over dependence upon a single, fallible method. Interviews and questionnaires intrude as a foreign element into the social setting they would describe, they create as well as measure attitudes, they elicit atypical roles and responses, they are limited to those who are accessible and will cooperate, and the responses obtained are produced in part by dimensions of individual differences irrelevant to the topic at hand. *But the principal objection is that they are used alone*. No research measure is without bias. Interviews and questionnaires must be supplemented by methods testing the same social science variables but having *different* methodological weaknesses.[1]

We will, then, examine several different ways of gathering data in the next chapter of this book. The important point for the moment, however, is the fact that different kinds of data, obtained in a variety of ways, can be brought to bear on our questions or hypotheses, and the kind of information we attempt to obtain plays a major part in shaping the plans of research:

☐ Surveys or questionnaires generally use "structured" responses to a large battery of questions and can be administered by relatively untrained people. They have the advantage of getting large quantities of comparable information at a relatively low cost from a large number of people, who do not need to be known intimately by the interviewers.

☐ Participant observation, in which the person who is collecting information observes and talks to and sometimes participates with the people he is studying as they go about their usual activities, allows greater intimacy and the observation of behavior--rather than verbalizations--over a prolonged period of time. The information obtained is frequently of a greater depth than is obtained by a questionnaire, and often involves the observer's interpretation of behavioral cues which are revealing but difficult to verify. In general, the number of people who can be studied is much smaller than in the case of surveys; and the information that is gathered, while rich in detail, is very dependent on the training and the sensitivity of the observer.

---

[1] Eugene J. Webb, *et al.*, *Unobtrusive Measures* (Chicago: Rand McNally and Company, 1966) p. 1.

☐ There is a good deal of debate in the social sciences about the relative virtues of surveys and participant observation. Much of it is unnecessary, if we remember the quotation above and recognize that all methods of collecting information--and all research designs--have their strengths and their weaknesses.

☐ Experiments in which extraneous factors are carefully controlled and the causal effects of a small number of variables are determined with relative precision are much admired and justly so. Experiments, however, are difficult to perform in the social sciences; and there are often ethical reasons for rejecting research designs which depend on them. At the same time, the logic underlying experiments is frequently employed implicitly or explicitly in much social science research and provides a model of the type of evidence in terms of standards of scientific rigor which we strive to approximate if not achieve exactly, as indicated in the reading by Samuel A. Stouffer.

III

The best research design, then, in our opinion, is one which tries to obtain information in a variety of ways, both to serve as a cross-check and to increase the scope of our information. This need for varied procedures is perhaps not so pressing in all fields of empirical inquiry, but it is of prime importance in studying social behavior for a number of reasons which are embedded in our subject matter.

First, the social sciences are concerned with the causes of social events and their consequences; and this involves us in a temporal sequence which demands more than the recollections of respondents (which may be quite distorted or incomplete) and their anticipations of the future (which may often be erroneous). A study executed at one point in time freezes the action when we really wish to examine the flow.

Second, the social sciences are deeply concerned with man's behavior and not just the words with which he clothes his actions and imperfectly reveals the contents of his mind. But at the same time, the meaning of behavior is difficult to discover simply by the use of overt physical signs. We must examine both words and actions, if we are to fully understand man as a social animal.

Third, the meaning of the behavior studied by the social sciences, and the causes and the consequences, are sometimes quite beyond the ken of the person exhibiting the behavior (the actor) either because there are unconscious motives at work or because there are other people involved who the actor knows imperfectly or not at all. The observer cannot rely merely on what the actor tells him, but must also observe the behavior of the people involved from many vantage points, either directly himself or through the accounts of others.

Fourth, the social sciences are seldom concerned with single cases but are largely concerned with collections of cases or units of observation--and they may be collections of individuals, social groups, populations, and so on. It is immediately obvious that the units which make up these various collections (or collectivities, as they are labelled in the jargon of sociology) are frequently dissimilar. Even when individuals or social groups or populations are identical with respect to some selected characteristics, they are very apt to differ with respect to other important characteristics. The result is that no unit in a particular collectivity will surely behave like all the other units under the same conditions or with the same set of antecedent events. And this means that we must study a number of cases or units to determine the average influence of a particular cause or the average size of a particular effect. (Since the collectivities studied by the social sciences are often quite large, we are generally forced to draw a sample; and this sample, as we will see later, must be representative of the collectivity if it is to be of much use.) We must examine the phenomenon we are interested in under different conditions, to see if the relationship we have hypothesized will still hold when the circumstances change.

All this adds up to the idea that at some point in our research, we will want to examine a fairly large number of cases, either in terms of data we have collected ourselves or of data collected by other researchers. And this inevitably requires elements of research design which are quite different from a full, detailed study of a handful of cases.

The three selections from the work of Professors Skolnick, Wolfgang, and Hunting and Neuwirth represent very different types of research design. The first stresses participant observation; the second emphasizes the use of official records; and the third depends primarily on survey techniques. None of these really makes a systematic effort to incorporate different types of data, it might be noted. Our argument for the combined use of various ways to obtain information clearly expresses an ideal of what research design should be like rather than a common practice. Aside from this, however, one of the most striking things about these descriptions of research projects is the extent to which the authors were forced to cut back from the wide-ranging scope of their original interest in designing a practical project. For example, Professors Hunting and Neuwirth start out with a general interest in the "litigating public." They end up with an examination of 164 people suffering mild injuries in automobile accidents in New York City in February, March, April, and May, 1957. Such a reduction is known intimately and sadly to all researchers in the social sciences as they attempt to fit their resources to their curiosity.

IV

One more issue should be noted, disconcerting though it may be. Given the complexities of social science research today, projects are seldom carried out by one person. Very often students and other assistants are involved and this means that the researcher finds himself cast in an administrative role. The danger here is that the person doing the research, but who is not in charge of it, may have very little real interest in the outcome or may in fact in some situations be extremely hostile. This is the problem which is discussed by Julius Roth in "Hired Hand Research." It is a difficulty which is not easily solved, but it is also an issue which can not be shoved under the carpet.

PLANNING A RESEARCH DESIGN

Research, as the authors of the following reading point out, is not carried out in a neat series of sequences beginning with a statement of purpose and concluding with a statement interpreting conclusions drawn from the data collected. There are what Professor Selltiz et al. call "additional activities" which must be undertaken if the researcher is not to find himself bogged down in a series of both theoretical and practical problems as a result of what amounts essentially to improper planning prior to the construction of hypotheses and collection of data. These "additional activities" Selltiz has divided into two categories: those related to the scientific requirements of the study and those related to the practical demands of the study.

In explaining what he means by activities related to the scientific requirements of a study, Selltiz refers us to a study carried out by Deutsh and Collins concerning interracial attitudes in public housing projects. The investigators, Selltiz tells us, were interested in the general problem of "relations between members of different . . . groups . . ." In particular, they wanted to observe interracial groups in face-to-face situations. Prior to any formulation of a specific research topic Deutsh and Collins engaged in preliminary research. They chose a setting--"large scale interracial housing"--and devoted eighteen months of systematic study to narrowing their area of inquiry and attempting to understand the dynamics of multi-racial public housing. The end result of their initial investigation was (1) the conclusion that a public housing setting would be a useful one for study of the effects of personal association between members of different racial groups; and (2) that their particular focus of inquiry would be directed at attempting to understand the effects of various occupancy patterns of the housing projects upon "social relations across racial lines, social standards for behavior with people of the other race, the general pattern of social relations in the project, and interracial attitudes." In other words, what Deutsh and Collins did was to systematically plan their research design by first choosing a setting and then observing the general problem of race relations within that milieu. Not to have done so, Selltiz is saying, would be to have left themselves open to the possibility of constructing an elaborate research design and then discovering not only that it dealt with too broad an area to produce useful data, but that the setting in which it was being carried out was not at all conducive to producing data relevant to the problem being studied. This is not to say that every research project should be preceded by a preliminary, investigatory study. What we are saying, however, is that a great deal of preliminary planning and study must precede the actual design of a research problem.

As for the "more practical" problems, Selltiz lists such things as budget planning, obtaining and administering funds, and acquiring and allocating skilled personnel. Referring to the Deutsh and Collins study, Selltiz shows how these more practical problems took the form of having to collect data mainly from white housewives, "because of the limited number of interviews that . . . resources made possible," recruiting skilled interviewers, and obtaining permission to interview in the projects used from both the New York and New Jersey public housing authorities. All this is to say that in planning and carrying out empirical research one will necessarily have to engage in administrative as well as scientific endeavors.

There is, however, one additional point which we feel is worth noting about the Deutsh and Collins study. It was designed with a social purpose in mind. They state, "Our decision to investigate the effects of the occupancy pattern . . . was not a matter of whim . . . . Unlike many of the factors in a housing project the occupancy pattern is directly determined by an administrative decision, . . . which is responsive . . . to the social climate . . . . These decisions are particularly apt to be influenced by knowledge of the consequences of the different patterns for tenant relations." To law students, we feel, this approach to empirical inquiry is most significant. It suggests that social science methods can be used in an advocate's context to bring about social and political change.

This material has been reproduced from Research Methods in Social Relations by Claire Selltiz, Marie Jahoda, Morton Deutsch, and Stuart W. Cook. Copyright 1951, (c) 1959 by Holt,Rinehart and Winston, Inc. No part of this material may be reproduced in any way without the written permission of Holt, Rinehart and Winston, Inc. pp. 8-23 Reprinted by permission.

Claire Selltiz, Research Methods in Social Relation
(New York: Holt, Rinehart, and Winston, 1964),
pp. 8-23.

Major Steps in Research

The object of this book is to describe in detail the procedures necessary to discover answers to questions through research. But since concern with detail often obscures perception of the whole, it is well, before embarking on the examination of specific procedures, to point out some over-all aspects of the research process.

The research process consists of a number of closely related activities that overlap continuously rather than following a strictly prescribed sequence. So interdependent are these activities that the first step of a research project largely determines the nature of the last. If subsequent procedures have not been taken into account in the early stages, serious difficulties may arise and prevent the completion of a study. Frequently these difficulties cannot be remedied at the time they become apparent because they are rooted in the earlier procedures. They can be avoided only by keeping in mind, at each step of the research process, the requirements of subsequent steps.

To be sure, as research proceeds from the conception of a theme for a study through the gathering of data to the production of a report and the application of the findings, the focus of attention will necessarily shift from one activity to the next. This shift reflects a difference in emphasis, however, rather than an exclusive concentration on one step. A mechanically consecutive sequence of procedures, in which one research step is entirely completed before the next is begun, is rarely, if ever, the experience of social scientists.

The usual pattern of *reporting* research creates an oversimplified expectation of what is involved in *doing* research. Customarily, a report on completed research, when it appears as an article in a technical journal, resembles, with minor modifications, the following model:

1. A statement of purpose is made in the form of *formulating the problem;*
2. A description of the *study design* is given;
3. The *methods of data collection* are specified;
4. The *results* are presented;
5. Frequently, there follows a section on *conclusions and interpretation.*

Whatever the individual variations from this model, published research strongly suggests the existence of a prescribed sequence of procedures, each step presupposing the completion of the preceding one. Although this model is entirely justified in the interest of economy of scientific reporting, it must not be mistaken for a model of the research process, which differs from it in two respects: (1) The research process almost never follows the neat sequential pattern of activities suggested in the organization of research reports; and (2) the process involves many additional activities which are rarely mentioned in published studies.

Some of these additional activities are related to the scientific requirements of the study; others to its practical demands. The apparently simple reporting of the methods of data collection, for example, summarizes decisions about the kinds of data needed and the most efficient way of collecting them, and the activities carried out in the development and pretesting of the data-collection instruments. In addition to these steps, related to the scientific requirements of the study, there are other, more "practical," demands: the budget must be planned; funds must be obtained and administered; personnel must be allocated and, in some cases, specially trained; the setting within which the data are to be collected must be explored and the cooperation of the people in it must be gained; etc. In addition, if the study is one designed to solve an immediate, practical problem, the anticipated application of the findings must be considered from the outset.

An Illustration

The manner in which each step influences, and is influenced by, others is perhaps best demonstrated by a brief case history of a research project. As an illustration, we shall use a study of interracial housing projects carried out by Deutsch and Collins (1951).* To a considerable extent we shall use the authors' own words, as they appear in the report of the study. But since, like all reports, this one suggests that the study was carried out in a neat series of separate steps, we shall intersperse comments about activities carried on in the course of the study which are not mentioned in the formal report. The account of this study will give an over-all view of the research process; each of the steps within the process will be discussed in detail in later chapters.

The investigators were members of a research group that had as a primary concern the study of relations between members of different racial and religious groups within the United States. As a further specification of its general area of interest, the group had chosen to concentrate on the study of situations in which members of different groups are in face-to-face contact. Clearly, these decisions mapped out a very general area of interest; the focus had to be specified much more sharply before a study could be undertaken.

In this case, the next step was somewhat atypical; it was the selection of a general *setting* for the research, even before the problem to be studied had been identified. This setting was to be large-scale interracial housing. Since no member of the research group had any real familiarity with such housing, the group had little basis for selecting a specific problem to be studied within that setting. Accordingly, they drew up a proposal which called for research to be carried out in two stages: (1) six months to be devoted to an exploratory study, consisting of interviews with housing experts, on the basis of which a specific question would be selected for more systematic study; (2) a year to be spent in a systematic study of the research question selected on the basis of the exploratory study. A foundation interested in research in the field of intergroup relations granted them the necessary funds.

One member of the research team took primary responsibility for the exploratory study. Preliminary work involved reading material about public and private interracial housing, compiling a list of housing experts whose experience and insights might make a valuable contribution, and arranging to interview them. In conferences with the director of the research group and with other colleagues, plans for

*Deutsch, M., and M.E. Collins, 1951. *Interracial housing: a psychological evaluation of a social experiment*. University of Minnesota.

these interviews were drawn up. Next, the investigator visited housing projects and housing officials in the eastern, midwestern, and far-western United States. Forty-two interviews, with managers of housing projects, race relations officials of public housing agencies, and other persons with relevant experience in housing, were carried out. The interviews lasted from two to eight hours. They covered the respondents' views about factors important in influencing relations between Negro and white tenants, problems on which research was especially needed, and suggestions about the feasibility of research on various problems.

These interviews strengthened the belief that the housing setting would be a useful one for study of the effects of personal association between members of different racial groups, pointed to the choice of public housing because of the scarcity of private interracial housing projects, and suggested a number of research questions, one of which the investigators selected for the systematic study. They explain both their choice of setting and their choice of a specific question on grounds of implications for social action, theoretical interest, and opportunities for research:

8

> Public housing has existed for more than ten years. It has had a chance to develop standard patterns and variations with respect to racial occupancy. The common pattern is complete segregation—Negroes and whites live in separate housing projects—but there are important exceptions. These exceptions and the variations among them, in effect, provide a natural social experiment which permits those engaged in carefully controlled social research to gather valuable information about conditions which make for wholesome race relations. Such research can, by providing scientifically grounded knowledge in the place of current uncertainties, aid policy makers in their efforts to live up to the principles of the democratic ethos in their official functioning.
>
> . . .
>
> Not only may such research be socially useful . . . but it might also have theoretical significance. Most of the studies of attempts to change prejudice have, by and large, been limited to influences which were relatively minor in relation to other

influences in the subject's social milieu. . . . there is reason to believe that the housing community . . . provides one of the few opportunities for interracial contact of a sufficiently intimate and extended nature to result in large attitudinal change. In light of the crudity of most attitude measurement techniques, changes of relatively large magnitude are necessary to make possible identification and measurement of significant theoretical relationships.

. . .

With a census of problems and issues in the field of interracial housing before us, our task was to select a problem for more definitive research. Many studies were suggested by our survey [of housing officials]: the effects of different policy decisions, the effects of different management procedures, the effects of different kinds of tenant activities. All of these and many more would be useful. From these possibilities, we selected for study the impact of different occupancy patterns: *the integrated interracial* pattern (families are assigned to apartments without consideration of race) and the *segregated bi-racial* pattern (Negro and white families live in the same project but are assigned to different buildings or to different parts of the project).

We make no claim that the occupancy pattern is the only important influence on racial relations in projects that house both white and Negro families. Quite on the contrary, our survey indicated that the state of racial relations in a project would be affected by many factors . . .: the neighborhood in which the project is located, the racial composition of the tenants, the attitudes of the management staff, project facilities, etc. The effects of the occupancy pattern would, of necessity, be colored by the influence of these other factors. . . .

Yet our decision to investigate the effects of the occupancy pattern, as our first systematic study in this area, was not a matter of whim. We had several reasons for the choice. From our survey of housing officials and from our theoretical expectations . . . it was apparent that the occupancy pattern would very probably prove to be one of the most crucial influences on race relations in housing projects. . . .

It is not often that social research executed without the

> instigation of an administrator will be useful in affecting administrative decisions. Nevertheless, a further important reason for our interest in the occupancy pattern is our belief that, under the present circumstances, *something can be done about it* and that research can offer guidance in the doing. Unlike many of the factors which affect the state of race relations in a housing project, the occupancy pattern is directly determined by an administrative decision, a decision which is responsive primarily to the "social climate" rather than to economic and physical limitations. . . . With the political atmosphere such as it is, with the alternative occupancy policies each having its pros and cons with respect to political feasibility, these decisions are particularly apt to be influenced by knowledge of the consequences of the different patterns for tenant relations.

Actually, it had not seemed feasible within the exploratory study to examine in detail the practical possibilities for research on each of the topics suggested; the investigators therefore did some further "scouting" before they settled on the effects of occupancy patterns as the focus of the second stage of the research. The purpose of this scouting was to determine whether it would be possible to find appropriate projects in which to carry out research on this topic and to secure permission from housing officials to do so. Preliminary inquiries suggested that it would be feasible; accordingly, the decision was made to study the effects of occupancy patterns.

For much of the next fourteen months, both investigators worked on the study full time. Their efforts were supplemented from time to time by conferences with a consultant and with colleagues, by a staff of interviewers, a group of coders, an organization which specialized in the machine processing of data, a secretary. But before these more specialized services were needed, it was necessary to formulate the research question more precisely, and to consider its relation to existing knowledge.

> Our statement of the problem . . . needed much additional formulation before it could serve as a guide to research. In effect, we had to develop hypotheses about the possible effects of the occupancy pattern from our knowledge of the essential differences between the two types of project we were

> studying and from a knowledge of basic socio-psychological principles. Otherwise research efforts would be dissipated on the investigation of factors not likely to be related to the occupancy pattern.
>
> . . . In the development of hypotheses, it is always simpler to deal with imaginary, idealized phenomena rather than with events as they occur in the real world. So, for the sake of convenience, let us conjure up two projects exactly alike except for their occupancy pattern. In one project, Negro and white families are assigned to apartments without consideration of race; in the other project, though Negro and white families live in the same project, they are assigned to different buildings or to different parts of the project. . . .
>
> From the point of view of race relations, what . . . are the essential differences between the two projects? . . . it seems to us that the two types of projects differ mainly with respect to (1) the physical and functional proximity of Negro and white families, (2) the social norms regarding racial relations implicit in the policy decision of the occupancy pattern by an official public authority, and (3) the relationship of the project to the broader community.
>
> These are the differences, but what are their effects? The answer to this question requires a further specification of interest. Effects upon what? Our original interest directs us to inquire about the effects of the occupancy pattern upon (1) social relations across racial lines, (2) the social standards for behavior with people of the other race, (3) the general pattern of social relations in the project, and (4) interracial attitudes.

As it became clear what the research question was to be and what type of evidence would be relevant to the answer, the investigators began to consider how to collect this evidence most economically and with least chance of being led to an incorrect conclusion—i.e., they began to develop their research design. The essential feature of the design was a comparison of the responses of residents in the two types of project. The development of a satisfactory research design is likely to be particularly difficult when the research is to be carried out in a real-life setting. In the case of this study, it required that the two types of project be alike in all relevant respects except occupancy pattern, and that the tenants in the two projects should have comparable initial attitudes. It was not easy to satisfy these conditions.

> In the abstract, the research design called for by the hypotheses is relatively uncomplicated. It simply requires a comparative study of a number of segregated and integrated interracial housing projects which are equated in all relevant respects other than the occupancy pattern. Practically speaking, however, the phrase "equated in all relevant respects" introduces enormous complications and difficulties. As any housing administrator will point out, "No two housing projects are alike."
>
> . . .
>
> We designed our study to overcome as many as possible of the difficulties which would otherwise distort our findings. First of all, we carefully selected the segregated and integrated interracial projects we were to study so they were as equivalent as possible in all relevant respects other than the occupancy pattern. Second, we decided not to limit our study to one project of each type; we stretched our funds so as to study two of each kind. And third, in our method of investigation we carefully collected data about factors other than the occupancy pattern, to determine whether our results could be explained in terms of these other factors.

Compromises with the ideal of projects "equated in all other relevant respects" had to be made. In the course of the preliminary scouting, it had been learned that the number of integrated projects was limited. At the time of the study, there were some fifty-odd cities throughout the United States that had public housing projects officially described as integrated; however, in most of these projects the great majority (over 90 per cent) of the families were of one race. It seemed reasonable to suppose that the effects of occupancy pattern would show up more clearly in projects where the numbers of white and Negro families were more evenly balanced. But such projects existed in less than ten cities. While the research question was being sharpened and the design of the study worked out, the search for appropriate projects continued. It was finally decided that New York City presented the best setting for selection of integrated projects; neighboring Newark provided matching segregated bi-racial projects. But interviewing could not be conducted in the projects without the consent of the housing officials. Although officials in both cities were interested and cooperative, several conferences were needed to complete arrangements with them.

The fact that the integrated and the segregated projects were in different cities introduced a problem; so too did certain characteristics of the projects. New York and Newark are different. Since the research funds did not provide for studying the attitudes of a sample of citizens of each city not living in public housing projects, it was not possible to work out a study design that would make it possible to judge to what extent observed differences between tenants in the two types of project might reflect differences in atmosphere in the two cities rather than the impact of occupancy pattern. Further, it developed that the projects that could best be matched on other grounds all had a high proportion of Negro tenants—from 40 to 70 per cent. This meant that the findings could properly be generalized only to projects with a similarly high proportion of Negroes. Finally, the projects in Newark were of a type known as area-segregated, with Negro and white families living in different sections of the project. Although this had the advantage of giving a clear-cut pattern of segregation, it meant that the findings might not hold for projects with less extreme patterns of segregation, such as those in which separate buildings occupied by Negro families and by white families are scattered throughout the project.

During the same period that the projects were being selected, the staff considered the selection of subjects within the projects and the method of collecting data. Here again there were limitations and compromises:

> From our preliminary survey and the theoretical analysis, it became clear that the focus of the study would be the tenants themselves. Since there was, by and large, little evidence of much organized or observable group activity, it was evident that the main source of our information must come through interviewing.
>
> . . .
>
> Because of the limited number of interviews that our resources made possible, it seemed advisable to concentrate on some segment of the tenant population. We decided to collect

our data primarily from white housewives. We made this choice mainly on the ground that the home is largely the domain of the woman. She spends more time in it than anyone else; she is, by and large, the initiator of activities and contacts that develop directly out of the home. Whether or not she "wears the pants in the family," she is the key person in activities centered in the place of residence.

It was not financially feasible to interview both Negro and white housewives in equal proportion. We decided to interview more white housewives as a result of our conviction that prejudiced interracial attitudes are more socially crucial among whites than among Negroes. Segregation and discrimination are, after all, enforced by the white, and not by the Negro segment of the population.

A more serious problem was that of determining whether the white housewives in the two types of project had had similar attitudes toward Negroes before they moved into the projects. Since all the projects had been occupied for several years, it was impossible to get measures of attitudes the tenants had held before moving in. Instead, the investigators drew on their knowledge of other studies, which had shown that such characteristics as religion, political attitudes, education, and previous experiences with Negroes are frequently related to attitudes toward Negroes. Accordingly, they included in the interview schedule questions on these points. If it turned out that the women in the integrated and in the segregated projects were similar in these respects, there would be some reassurance—though by no means certainty—that they were similar in their initial attitudes toward Negroes. Unfortunately, it developed that there were differences between the two groups in religion, political attitudes, and education; thus these factors had to be taken into account throughout the analysis.

Decisions such as this, about what information was needed, led naturally to consideration of the interview schedule, the instrument by which the data were to be collected. The process of developing questions was simplified by the fact that much research on attitudes toward Negroes had already been carried out. The interview as finally developed covered five major areas: the attitudes of the housewives toward living in the project, attitudes toward Negroes, the amount and intimacy of contact with other women in the project, the social supports for attitudes, and the characteristics of the housewives.

While the interview schedule was being developed, the samples of women to be interviewed were being drawn from the lists of tenants in the projects. During this same period, one of the staff members began taking steps to recruit the number of skilled interviewers who would be needed to collect the data within a reasonable time. The interviewers were graduate students in social work and psychology. As soon as the interview schedule was in a form that seemed reasonably satisfactory, a "pilot test" was carried out; two or three of the most experienced interviewers, and the investigators themselves, carried out interviews with a small number of white housewives in other housing projects similar to those that had been selected for the study. As was expected, these pilot interviews pointed up questions that were not clear, those that needed especially careful handling to avoid antagonizing the respondents, those that did not seem to elicit the information they were intended to get. Changes were made in the interview schedule to overcome these difficulties. After another set of pilot interviews had been checked, all interviewers were trained in use of the schedule. Each interviewer spent approximately twelve hours in training sessions and conducted two practice interviews with residents of projects not included in the study.

Finally, the actual interviewing got under way. Five hundred interviews were conducted: four hundred with white housewives, one hundred with Negro housewives. The interviews lasted, on the average, from an hour to an hour and a half. There were nineteen interviewers, and all the interviews were completed within a month. During the interviewing, the investigators spent time in each of the projects, supervising the assignments and inspecting the interviews as they were completed.

Once data collection had taken place according to the specified research design, a number of irreversible decisions had been made which largely determined the next step—analysis and interpretation. In this study, the plan of data collection made it possible to compare the white housewives in the area-segregated and the integrated projects in terms of: (1) their extent of association with Negroes, (2) their perception of the social norms concerning association with Negroes,

(3) the relation between perceived social norms and extent of association, (4) their attitudes toward Negroes living in the project and Negroes in general, and (5) their attitudes toward living in an interracial project.

The limitations and compromises that have already been pointed out meant that it was *not* possible to be entirely sure that the differences found between the housewives in the two types of project represented the effect of occupancy pattern rather than general differences in attitude in the two cities or differences in attitude between the two groups of white housewives that existed before they ever moved into the projects. Nor was it possible to say whether the findings would hold for projects with smaller proportions of Negroes and for projects where the pattern of segregation was less marked—for example, in projects segregated by buildings rather than total sections.

Finally, the pattern of area-segregation in the Newark projects meant that, even if it was found that the white housewives living in integrated projects had more favorable attitudes toward Negroes than those in the segregated projects, it would be difficult to draw inferences about the *processes* that contributed to the difference in attitudes. In thinking about the ways in which different occupancy patterns might be expected to lead to differences in attitudes, the investigators considered two major factors: (1) the extent to which Negro and white tenants had occasion to see and to meet each other and thus had natural opportunities to become acquainted and (2) the implications of the integrated and segregated arrangements in terms of social approval or disapproval of association between white and Negro families. The pattern of segregation in the Newark projects, with white and Negro families living in separate sections of the project rather than in separate buildings scattered throughout the project, made it impossible to disentangle the effects of perceived social standards from those of simple physical proximity, since in these area-segregated projects not only did white and Negro families live relatively far from each other but the segregated arrangement suggested social disapproval of interracial association. This was a limitation primarily on the potential contribution of the study to theoretical knowledge of the dynamics of attitude change; it did not interfere with the possibility of gathering evidence on the practical question of the effects on attitudes of living in an integrated or an area-segregated project.

In order to provide answers to those questions which *could* be answered by a study having this particular design, it was necessary, of course, to analyze the replies to the interviews. The function of the analysis was to make possible a comparison of the women in the integrated projects with those in the segregated projects in terms of the possible effects (or *dependent variables*) which had been specified in the formulation of the research problem and in the construction of the interview schedule.

The first step in the analysis was to derive from the answers to the interview questions an indicator of each dependent variable with which the study was concerned, or sometimes several indicators. For example, in the general area of social relations within the projects, the main dependent variable with which the investigators were concerned was *intimacy of contact with Negroes* on the part of white housewives. The interview provided three indicators of this variable: (1) whether the housewife reported knowing any Negro residents "pretty well"; (2) whether she included at least one Negro among the five people in the project whom she knew best; and (3) the number of different types of neighborly contact she reported with Negro women in the project.

In the general area of interracial attitudes, the investigators distinguished eleven different variables, such as degree of esteem for Negroes in the project, degree of esteem for Negroes in general, degree of friendly feeling toward Negroes in general, etc. For each of these variables, one or more indicators were derived from the answers to the interview questions or from ratings by the interviewers.

The second step in the analysis was to prepare tables showing the distribution, on each index, of respondents in each of the four projects. These tables were then examined and subjected to *tests of statistical significance* to determine whether the differences between housewives in the segregated projects and housewives in the integrated projects on a particular variable were so large that they could not be reasonably interpreted as resulting simply from accidents in the random selection of subjects.

Almost all the tables showed large differences between housewives in the two types of project, and the differences were consistently in the direction of more social contacts with Negroes and more favorable attitudes toward Negroes in the integrated projects. Most of the differences were far too large to be reasonably attributed to sampling fluctuations.

The next step in the analysis was to investigate the distribution of respondents on each of the *background variables*, or characteristics of respondents thought to be related to attitudes toward Negroes. As we have previously noted, there were substantial differences between respondents in the two types of project in religion, political attitudes, and education. It was consequently necessary to prepare additional sets of tables showing, for each index, the scores of Protestant women in each of the four projects, of Catholic women, and of Jewish women. From these tables it was possible to determine that the previously found relationships between occupancy pattern and the dependent variables remained approximately the same when Protestant, Catholic, and Jewish women were examined separately—that is, when religion was "held constant." Similar tables were prepared "holding constant" education and political attitudes, with similar results.

On the basis of this analysis it was concluded that the differences in attitudes and social relations with Negroes between housewives in the two types of project could not be accounted for by the differences in religion, education, or political views. Similar analyses indicated that these differences could not be accounted for by differences in initial attitudes toward Negroes (as reported by the respondents themselves), or by differences in expectations regarding the occupancy pattern at the time of moving in. It was not possible to rule out the possible effect of differences in the "social climate" of New York City and Newark, or of other uncontrolled variables that were not tapped by the interview schedule; but it was tentatively concluded that the differences in the dependent variables were due to the differences in occupancy pattern in the two types of project.

Interest then turned to the question of *how* the difference in occupancy pattern might produce the observed differences in attitudes toward Negroes. The investigators had included in their interview schedule a question on the ways in which someone living in the project might get to know the Negro women in the project. The most common ways mentioned by housewives in the integrated projects were meeting the Negro women as neighbors in the building, or on benches outside the building. In the segregated projects only half as many women mentioned *any* ways in which one would be likely to get to know the Negro women, and the contacts that were most frequently mentioned, such as meeting in stores or in streets around the project, did not provide natural opportunities for extended conversation.

Another analysis was made comparing the housewives in the integrated projects who reported that their attitudes toward Negroes had become more favorable since living in the project with those who reported no change in their attitudes. This showed, among other things, that the women whose attitudes had changed tended to be those who had more intimate contact with the Negro women in the project and who believed that their white friends in the project approved of their friendly association with the Negro women. As was pointed out earlier, the study design did not make possible an evaluation of the relative influence of association with Negroes and of perceived social approval of such association on the process of attitude change.

Because of this and other uncertainties of interpretation mentioned earlier, plans were made almost immediately for another study, which would build on this one and carry its findings still further (Wilner, Walkley, and Cook, 1955).*This new study was carried out in four cities (all of them outside the New York metropolitan area). The projects all had a quite small proportion of Negroes, and the segregated projects took the form of scattered Negro and white buildings rather than separate areas. These two latter characteristics made it possible to examine separately the influence of physical proximity and of implied official standards, since some white women in the building-segregated projects lived closer to Negroes than did some white women in the integrated projects. The findings of this study indicated that physical proximity was the more important influence.

This illustration can indicate only in barest outline the nature of the research process. As would be the case with any illustration, it does not cover all the possibilities of interrelation and interdependence of research steps. The pattern of interaction among the various procedures that constitute a scientific inquiry will, of course, vary from study to study. The point of this illustration is to show not only that early

Wilner, D.M., R.P. Walkley, and S.W. Cook, 1955. Human relations in interracial housing: a study of the contact hypothesis. University of Minnesota.

steps influence subsequent ones—an obvious matter—but also that the interaction of each step with others is a major consideration in its selection, and that subsequent steps often lead to a reconsideration of preceding ones. Social research is not a deductive process, in which everything follows from some clearly defined premises; it is a continuous search for truth, in which tentative answers lead to a refinement of the questions to which they apply and of the procedures by which they were obtained.

THE MODEL OF EXPERIMENTAL DESIGN

Samuel Stouffer, in the following reading, suggests that part of the reason for social science not fulfilling the prophecy that it would produce "dazzling miracles" parallel to those produced by the natural sciences is because it has been afflicted with what he calls the basic problem of the thoughtways of our culture--"the implicit assumption that anybody with a little common sense and a few facts can come up at once with the correct answer in any subject."

An example of what Stouffer means is best given by referring to his discussion of the controlled experiment. Briefly, a controlled experiment is "one where the researcher is essentially thinking in terms of cause and effect. That is, he has in mind one or more variable, variations in which can be used to explain variation in another variable." (These causal dimensions are termed independent variables and the values to be explained are called variables.) A control group, then, which is not subjected to the independent variable, becomes critical to a fundamental understanding of how and to what degree the independent variable really brings about the effect predicted by the particular hypothesis being tested. For example, in testing the hypothesis that indoctrination movies would boost the morale of enlisted men during World War II, Stouffer interviewed two groups of men--one group had seen the films, the other had not. To have proceeded without a control group (a group which had not seen the films) would have resulted in the erroneous but apparently reasonable assertion that the films boost morale, since those who had seen the films had high morale, since those not seeing the films had equally high morale. As Stouffer notes, "The test of whether a difference d is attributable to what we think it is attributable to is whether d is significantly larger than d' [a control group]."

There is, however, more to the Stouffer article than just a discussion of the benefits of a controlled experiment. He is saying that too much contemporary research is trivial, that social scientists have not devoted themselves to building a modern theory of society, that they are engaging in study that is "not really important." We interpret his criticism of much empirical research as being directed at not only research which has no social relevance, but research which is undertaken because the particular problem allows the researcher to construct a relatively simple research design which produces the obvious rather than scientific understanding.

From Samuel Stouffer, "Some Observations on Study Design," The American Journal of Sociology, 55 (January, 1950), pp. 356-359. Published by the University of Chicago Press. Reprinted by permission.

Samuel A. Stouffer, "Some Observations on Study Design," The American Journal of Sociology, 55 (January; 1950), pp. 355-361.

As a youth I read a series of vigorous essays in the *Century Magazine* by its editor, the late Glenn Frank. His theme was that the natural sciences had remade the face of the earth; now had arrived the age of the social sciences. The same techniques which had worked their miracles in physics, chemistry, and biology should, in competent hands, achieve equally dazzling miracles in economics, political science, and sociology. That was a long time ago. The disconcerting fact is that people are writing essays just like that today. Of course, the last two decades have seen considerable progress in social science—in theory, in technique, and in the accumulation of data. It is true that the number of practitioners is pitifully few; only a few hundred research studies are reported annually in sociology, for example, as compared with more than twenty thousand studies summarized annually in *Biological Abstracts*. But the bright promise of the period when Frank was writing has not been fulfilled.

Two of the most common reasons alleged for slow progress are cogent, indeed.

The data of social science are awfully complex, it is said. And they involve values which sometimes put a strain on the objectivity of the investigator even when they do not incur resistance from the vested interests of our society. However, an important part of the trouble has very little to do with the subject matter of social science as such but, rather, is a product of our own bad work habits. That is why this paper on the subject of study design may be relevant. So much has been spoken and written on this topic that I make no pretense to originality. But in the course of a little experience, especially in an effort during the war to apply social psychology to military problems, and in an undertaking to nurture a new program of research in my university, I have encountered some frustrations which perhaps can be examined with profit.

A basic problem—perhaps *the* basic problem—lies deeply imbedded in the thoughtways of our culture. This is the implicit assumption that anybody with a little common sense and a few facts can come up at once with the correct answer on any subject. Thus the newspaper editor or columnist, faced with a column of empty space to fill with readable English in an hour, can speak with finality and authority on any social topic, however complex. He might not attempt to diagnose what is wrong with his sick cat; he would call a veterinarian. But he knows precisely what is wrong with any social institution and the remedies.

In a society which rewards quick and confident answers and does not worry about how the answers are arrived at, the social scientist is hardly to be blamed if he conforms to the norms. Hence, much social science is merely rather dull and obscure journalism; a few data and a lot of "interpretation." The fact that the so-called "interpretation" bears little or no relation to the data is often obscured by academic jargon. If the stuff is hard to read, it has a chance of being acclaimed as profound. The rewards are for the answers, however tediously expressed, and not for rigorously marshaled evidence.

8

In the army no one would think of adopting a new type of weapon without trying it out exhaustively on the firing range. But a new idea about handling personnel fared very differently. The last thing anybody ever thought about was trying out the idea experimentally. I recall several times when we had schemes for running an experimental tryout of an idea in the sociopsychological field. Usually one of two things would happen: the idea would be rejected as stupid without a tryout (it may have been stupid, too) or it would be seized on and applied generally and at once. When the provost marshal wanted us to look into the very low morale of the MP's, our attitude surveys suggested that there was room for very much better selectivity in job assignment. There were routine jobs like guarding prisoners which could be given to the duller MP's, and there were a good many jobs calling for intelligence, discretion, and skill in public relations. We thought that the smarter men might be assigned to these jobs and that the prestige of these jobs would be raised further if a sprinkling of returned veterans with plenty of ribbons and no current assignment could be included among them. We proposed a trial program of a reassignment system in a dozen MP outfits for the purpose of comparing the resulting morale with that in a dozen matched outfits which were left untouched. Did we get anywhere? No. Instead, several of our ideas were put into effect immediately throughout the army without any prior testing at all.

The army cannot be blamed for behavior like that. In social relations it is not the habit in our culture to demand evidence for an idea; plausibility is enough.

To alter the folkways, social science itself must take the initiative. We must be clear in our own minds what proof consists of, and we must, if possible, provide dramatic examples of the advantages of relying on something more than plausibility. And the heart of our problem lies in study design *in advance*, such that the evidence is not capable of a dozen alternative interpretations.

Basically, I think it is essential that we always keep in mind the model of a controlled experiment, even if in practice we may have to deviate from an ideal model. Take the simple accompanying diagram.

	Before	After	After − Before
Experimental group	x_1	x_2	$d = x_2 - x_1$
Control group	x_1'	x_2'	$d' = x_2' - x_1'$

The test of whether a difference d is attributable to what we think it is attributable to is whether d is significantly larger than d'.

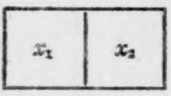
We used this model over and over again during the war to measure the effectiveness of orientation films in changing soldiers' attitudes. These experiences are described in Volume III of our *Studies in Social Psychology in World War II.*[1]

One of the troubles with using this careful design was that the effectiveness of a single film when thus measured turned out to be so slight. If, instead of using the complete experimental design, we simply took an unselected sample of men and compared the attitudes of those who said they had seen a film with those who said they had not, we got much more impressive differences. This was more rewarding to us, too, for the management wanted to believe the films were powerful medicine. The gimmick was the selective fallibility of memory. Men who correctly remembered seeing the films were likely to be those most sensitized to their message. Men who were bored or indifferent may have actually seen them but slept through them or just forgot.

Most of the time we are not able or not patient enough to design studies containing all four cells as in the diagram above. Sometimes we have only the top two cells, as in the accompanying diagram. In this situation

---

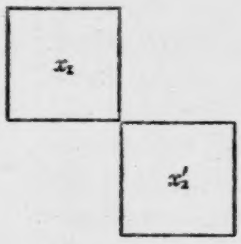
[1] Carl I. Hovland, Arthur A. Lumsdaine, and Fred D. Sheffield, *Experiments in Mass Communication* (Princeton: Princeton University Press, 1949).



$d = x_2 - x_1$

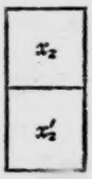
we have two observations of the same individuals or groups taken at different times. This is often a very useful design. In the army, for example, we would take a group of recruits, ascertain their attitudes, and restudy the same men later. From this we could tell whose attitudes changed and in what direction (it was almost always for the worse, which did not endear us to the army!). But exactly what factors in the early training period were most responsible for deterioration of attitudes could only be inferred indirectly.

The panel study is usually more informative than a more frequent design, which might be pictured thus:



Here at one point in time we have one sample, and at a later point in time we have another sample. We observe that our measure, say, the mean, is greater for the recent sample than for the earlier one. But we are precluded from observing which men or what type of men shifted. Moreover, there is always the disturbing possibility that the populations in our two samples were initially different; hence the differences might not be attributable to conditions taking place in the time interval between the two observations. Thus we would study a group of soldiers in the United States and later ask the same questions of a group of soldiers overseas. Having matched the two groups of men carefully by branch of service, length of time in the army, rank, etc., we hoped that the results of the study would approximate what would be found if the same men could have been studied twice. But this could be no more than a hope. Some important factors could not be adequately controlled, for example, physical conditions. Men who went overseas were initially in better shape on the average than men who had been kept behind; but, if the follow-up study was in the tropics, there was a chance that unfavorable climate already had begun to take its toll. And so it went. How much men overseas changed called for a panel study as a minimum if we were to have much confidence in the findings.

A very common attempt to get the results of a controlled experiment without paying the price is with the design that might be as shown in the accompanying diagram. This



is usually what we get with correlation analysis. We have two or more groups of men whom we study at the same point in time. Thus we have men in the infantry and men in the air corps and compare their attitudes. How much of the difference between x'_2 and x_2 we can attribute to experience in a given branch of service and how much is a function of attributes of the men selected for each branch we cannot know assuredly. True, we can try to rule out various possibilities by matching; we can compare men from the two branches with the same age and education, for example. But there is all too often a wide-open gate through which other uncontrolled variables can march.

Sometimes, believe it or not, we have only one cell:

When this happens, we do not know much of anything. But we can still fill pages of social science journals with "brilliant analysis" if we use plausible conjecture in supplying missing cells from our imagination. Thus we may find that the adolescent today has wild ideas and conclude that society is going to the dogs. We fill in the dotted cell representing our own yesterdays with hypothetical data, where x_1 represents us and x_2 our offspring.

x_1	x_2

The tragicomic part is that most of the public, including, I fear, many social scientists, are so acculturated that they ask for no better data.

I do not intend to disparage all research not conforming to the canons of the controlled experiment. I think that we will see more of full experimental design in sociology and social psychology in the future than in the past. But I am well aware of the practical difficulties of its execution, and I know that there are numberless important situations in which it is not feasible at all. What I am arguing for is awareness of the limitations of a design in which crucial cells are missing.

Sometimes by forethought and patchwork we can get approximations which are useful if we are careful to avoid overinterpretation. Let me cite an example:

In Europe during the war the army tested the idea of putting an entire platoon of Negro soldiers into a white infantry outfit. This was done in several companies. The Negroes fought beside white soldiers. After several months we were asked to find out what the white troops thought about the innovation. We found that only 7 per cent of the white soldiers in companies with Negro platoons said that they disliked the idea very much, whereas 62 per cent of the white soldiers in divisions without Negro troops said they would dislike the idea very much if it were tried in their outfits. We have:

	Before	After
Experimental		7%
Control		62%

Now, were these white soldiers who fought beside Negroes men who were naturally more favorable to Negroes than the cross-section of white infantrymen? We did not think so, since, for example, they contained about the same proportion of southerners. The point was of some importance, however, if we were to make the inference that actual experience with Negroes reduced hostility from 62 to 7 per cent. As a second-best substitute, we asked the white soldiers in companies with Negro platoons if they could recall how they felt when the innovation was first proposed. It happens that 67 per cent said they were initially opposed to the idea. Thus we could tentatively fill in a missing cell and conclude that, under the conditions obtaining, there probably had been a marked change in attitude.

Even if this had been a perfectly controlled experiment, there was still plenty of chance to draw erroneous inferences. The conclusions apply only to situations closely approximating those of the study. It happens, for example, that the Negroes involved were men who volunteered to leave rear-area jobs for combat duty. If other Negroes had been involved, the situation might have been different. Moreover, they had white officers. One army colonel who saw this study and whom I expected to ridicule it because he usually opposed innovations, surprised me by offering congratulations. "This proves," he said, "what I have been arguing in all my thirty years in the army—that niggers will do all right if you give 'em white officers!" Moreover, the study applied only to combat experience. Other studies would be needed to justify extending the findings to noncombat or garrison duty. In other words, one lone study, however well designed, can be a very dangerous thing if it is exploited beyond its immediate implications.

Now experiments take time and money, and there is no use denying that we in social science cannot be as prodigal with the replications as the biologist who can run a hundred experiments simultaneously by growing plants in all kinds of soils and conditions. The relative ease of experimentation in much—not all—of natural science goes far to account for the difference in quality of proof demanded by physical and biological sciences, on the one hand, and social scientists, on the other.

Though we cannot always design neat experiments when we want to, we can at least keep the experimental model in front of our eyes and behave cautiously when we fill in missing cells with dotted lines. But there is a further and even more important operation we can perform in the interest of economy. That lies in our choice of the initial problem.

Professor W. F. Ogburn always told his students to apply to a reported research conclusion the test, "How do you know it?" To this wise advice I should like to add a further question: "What of it?" I suspect that if before designing a study we asked ourselves, more conscientiously than we do, whether or not the study really is important, we would economize our energies for the few studies which are worth the expense and trouble of the kind of design I have been discussing.

Can anything be said about guides for selecting problems? I certainly think so. That is where theory comes in and where we social scientists have gone woefully astray.

Theory has not often been designed with research operations in mind. Theory as we have it in social science serves indispensably as a very broad frame of reference or general orientation. Thus modern theories of culture tell us that it is usually more profitable to focus on the learning process and the content of what is learned rather than on innate or hereditary traits. But they do not provide us with sets of interrelated propositions which can be put in the form: If x_1, given x_2 and x_3, then there is strong probability that we get x_4. Most of our propositions of that form, sometimes called "theory," are likely to be *ad hoc* common-sense observations which are not deducible from more general considerations and which are of the same quality as the observation, "If you stick your hand in a fire and hold it there, you will get burned."

Now in view of the tremendous cost in time and money of the ideal kind of strict empirical research operations, it is obvious that we cannot afford the luxury of conducting them as isolated fact-finding enterprises. Each should seek to be some sort of *experimentum crucis*, and, with rare exceptions, that will only happen if we see its place *beforehand* in a more general scheme of things. Especially, we need to look for situations where two equally plausible hypotheses deducible from more general theory lead to the expectation of different consequences. Then, if our evidence supports one and knocks out the other, we have accomplished something.

The best work of this sort in our field is probably being done today in laboratory studies of learning and of perception. I do not know of very good sociological examples. Yet in sociology experiments are possible. One of the most exciting, for example, was that initiated long before the war by Shaw and McKay to see whether co-operative effort by adult role models within a delinquent neighborhood would reduce juvenile delinquency. So many variables are involved in a single study like that that it is not easy to determine which were crucial. But there was theory behind the study, and the experimental design provided for controlling at least some variables.

It may be that in sociology we will need much more thinking and many more descriptive studies involving random ratlike movements on the part of the researcher before we can even begin to state our problems so that they are in decent shape for fitting into an ideal design. However, I think that we can reduce to some extent the waste motion of the exploratory period if we try to act as if we have some a priori ideas and keep our eyes on the possible relevance of data to these ideas. This is easier said than done. So many interesting rabbit tracks are likely to be uncovered in the exploratory stages of research that one is tempted to chase rabbits all over the woods and forget what his initial quarry was.

Exploratory research is of necessity fumbling, but I think that the waste motion can be reduced by the self-denying ordinance of deliberately limiting ourselves to a few variables at a time. Recently two of my colleagues and myself have been doing a little exploratory work on a problem in the general area of social mobility. We started by tabulating some school records of fifty boys in the ninth grade of one junior high school and then having members of our seminar conduct three or four interviews with each boy and his parents. We had all the interviews written up in detail, and we had enough data to fill a book—with rather interesting reading, too. But it was a very wasteful process because there were just too many intriguing ideas. We took a couple of ideas which were deducible from current general theory and tried to make some simple fourfold tables. It was obvious that, with a dozen variables uncontrolled, such tables meant little or nothing. But that led us to a second step. Now we are trying to collect school records and a short questionnaire on two thousand boys. We will not interview all these boys and their parents in detail. But, with two thousand cases to start with, we hope to take a variable in which we are interested and find fifty boys who are plus on it and fifty who are minus, yet who are approximately alike on a lot of other things. A table based on such matched comparisons should be relatively unambiguous. We can take off from there and interview those selected cases intensively to push further our exploration of the nexus between theory and observation. This, we think, will be economical, though still exploratory. Experimental manipulation is far in the future in our problem, but we do hope we can conclude the first stage with a statement of some hypotheses susceptible to experimental verification.

I am not in the least deprecating exploratory work. But I do think that some orderliness is indicated even in the bright dawn of a youthful enterprise.

One reason why we are not more orderly in our exploratory work is that all too often what is missing is a sharp definition of a given variable, such that, if we wanted to take a number of cases and even throw them into a simple fourfold table, we could.

Suppose we are studying a problem in which one of the variables we are looking for is overprotection or overindulgence of a child by his mother. We have a number of case histories or questionnaires. Now how do we know whether we are sorting them according to this variable or not? The first step, it would seem, is to have some way of knowing whether we are sorting them along any single continuum, applying the same criteria to each case. But to know this we need to have built into the study the ingredients of a scale. Unless we have some such ingredients in our data, we are defeated from the start. This is why I think the new interest social scientists are taking in scaling techniques is so crucially important to progress. In particular, the latent-structure theory developed by Paul F. Lazarsfeld, which derives Louis Guttman's scale as an important special case, is likely to be exceedingly useful, for it offers criteria by which we can make a small amount of information go a long way in telling us the logical structure of a supposed variable we are eager to identify. The details of Guttman's and Lazarsfeld's work[2] are likely to promote a good deal of attack and controversy. Our hope is that this will stimulate others to think such problems out still better and thus make their work obsolete as rapidly as possible.

Trying to conduct a social science investigation without good criteria for knowing whether a particular variable may be treated as a single dimension is like trying to fly without a motor in the plane. Students of the history of invention point out that one reason why the airplane, whose properties had been pretty well thought out by Leonardo da Vinci, was so late in development was the unavailability of a lightweight power plant, which had to await the invention of the internal combustion motor. We are learning more and more how to make our light-weight motors in social science, and that augurs well for the future. But much work is ahead of us. In particular, we desperately need better projective techniques and better ways of getting respondents to reveal attitudes which are too emotionally charged to be accessible to direct questioning. Schemes like the latent-structure theory of Lazarsfeld should speed up the process of developing such tests.

---

[2] Samuel A. Stouffer, Louis Guttman, Edward A. Suchman, Paul F. Lazarsfeld, Shirley A. Star, and John A. Clausen, *Measurement and Prediction* (Princeton: Princeton University Press, 1949).

I have tried to set forth the model of the controlled experiment as an ideal to keep in the forefront of our minds even when by necessity some cells are missing from our design. I have also tried to suggest that more economy and orderliness are made possible, even in designing the exploratory stages of a piece of research—by using theory in advance to help us decide whether a particular inquiry would be important if we made it; by narrowing down the number of variables; and by making sure that we can classify our data along a particular continuum, even if only provisionally. And a central, brooding hope is that we will have the modesty to recognize the difference between a promising idea and proof.

Oh, how we need that modesty! The public expects us to deal with great problems like international peace, full employment, maximization of industrial efficiency. As pundits we can pronounce on such matters; as citizens we have a duty to be concerned with them; but as social scientists our greatest achievement now will be to provide a few small dramatic examples that hypotheses in our field can be stated operationally and tested crucially. And we will not accomplish that by spending most of our time writing or reading papers like this one. We will accomplish it best by rolling up our sleeves and working at the intricacies of design of studies which, though scientifically strategic, seem to laymen trivial compared with the global concerns of the atomic age. Thereby, and only thereby, I believe, can we some day have the thrilling sense of having contributed to the structure of a social science which is cumulative.

THE BUILDING OF THEORY

In introducing his study Justice Without Trial, Jerome Skolnick says, "Since we do not presently have a theory of the administration of criminal justice we cannot say whether such variables as community size, rural-urban character . . . political links of court and legislature . . . are to be equally or differently evaluated." What Professor Skolnick is referring to, of course, is the need for a body of comparable studies prior to any theory-building concerning behavior. He does not, however, draw any generalizations concerning the criminal justice system in his own study. He does not generalize simply because he cannot. At the time he engaged in his research, there was not a sufficient body of literature pertaining to the criminal process against which he could evaluate his findings. What he does do, and what is of significance to the student of social science research, is to clearly define and list the social, economic, and historical facts incident both to the police department under observation and to the city, "Westville," in which the research was undertaken. By doing this he has insured that his work can be used as part of a general body of literature which eventually will lead to the development of a theory of the administration of criminal justice.

The Skolnick reading, however, illustrates more than the need to be cognizant of building theory when engaging in empirical research. It gives us an insight into the development of a type of research design--participant observation. In undertaking his study, Skolnick tells us he adopted as his model William F. H. Whyte's study, Street Corner Society, which presents a view of a delinquent sub-culture as being both complex and in conflict with conventional morality. What Skolnick is suggesting here is that by becoming a participant observer one can reach a point of empathy with his subjects and thereby obtain a better perspective than can be acquired through the use of questionnaires, surveys, and interviews. While we find ourselves in agreement with Skolnick as to the value of participant observation, we would caution the student that the mere observation of a group will not produce the kind of revealing results which Skolnick cites in referring to Street Corner Society. As Professor Skolnick points out, "A participant observer . . . must genuinely be prepared to see the world through the eyes of his subjects. Such a position does not undermine scientific objectivity . . . rather scientific knowledge should be enhanced, provided the participant observer is neither captured or repelled by his subjects."

From Jerome H. Skolnick, *Justice Without Trial* (New York: John Wiley & Sons, 1967) pp. 6; 23-37. Reprinted by permission.

Jerome H. Skolnick, Justice Without Trial
(New York: John Wiley & Sons, 1967), pp.23-37.

THE purpose of this chapter is to answer standard methodological questions: the nature of the community being studied, the character of the police department, the time of the study, and the methods used to obtain information. Special emphasis is given to the developmental aspect of the research, in effect, to discussing the question of how a study initially considered as exploratory grows into a variety of research interests and assignments.

One way of presenting a methodological background is to present only results, suggesting, perhaps unintentionally, that topics were worked out systematically beforehand; but it is neither true nor fair to the method used here to suggest a greater degree of initial order than actually existed. A reader who finds this sort of report tiresome may prefer to skip the chapter, especially the section on phases of data collection and those following. On the other hand, the reader interested in questions of how relations with police and others in the criminal law community are developed; of the nature of participant observation in such a setting; of ethical problems in this research stance, should find these sections of the chapter interesting.

THE SETTING OF THE STUDY

Data for this book were drawn from a study of criminal law officials conducted by the author mainly in Westville, a city of approximately four hundred thousand with a nonwhite population of about 30 per cent. For the sake of broad comparison, two weeks were also spent in Eastville, a city of comparable size, nonwhite population, industry, and commerce. I also visited police departments and rode patrol in two other cities. In addition, federal, state, and local law enforcement personnel were interviewed at several lengthy conferences. All this, however, was background material to put Westville in perspective. In the main, therefore, this is a study of the administration of criminal justice in one middle-sized city.

How representative of the United States is Westville and its administration of criminal justice? There are fifty-one separate jurisdictions in the United States, each with a body of statutory and case law, although some, modeled on the Federal Rules of Criminal Procedure, are more alike than others. In addition, there are practical differences in the administration of justice from county to county and city to city in any state, and among federal districts. Furthermore, in assessing the representativeness of a specific jurisdiction, it is difficult to single out with assurance which criteria are most relevant to determining the legal character of a community. Since we do not presently have a theory of the administration of criminal justice, we cannot say whether such variables as community size, rural-urban character, regional characteristics, degree of industrialization, racial and ethnic composition, political links of court and legislature, or history of corruption are to be equally or differentially evaluated. Lacking such a theory, we are in no position to assert which of these "test" variables it is important to control. To point out the difficulty of estimating the representative quality of a criminal courts community, is not the same, however, as to suggest that comparability is not desirable. To be able to draw inferences about the etiology of distinctive patterns, several case studies along the lines of the one that follows would be preferable. Given limited time and resources it was not possible, however, to complete even two intensive case studies.

Theory or no, it may be assumed that the social, economic, and historic facts of the city studied, the general outlines of the environment, are relevant to the operation of the machinery of justice. Such factual background is found in Appendix A. Factual background about the police department is in Appendix B. In addition to these fundamentals, there are several outstanding characteristics of Westville and its administration of criminal law that are important enough to bring to immediate attention.

First, Westville is a "real" city. It is neither a college town where the role of a researcher might be limited by his connection with the university, nor a megalopolis like New York or London that dominates the surrounding country. In New Haven, for instance (where I lived for almost ten years), social research is inevitably affected by the history of town-gown relations. Professional relations with the New Haven police department tend to have a special edge of delicacy, since the university and public institutions in general are interested in improving relations with one another. By contrast, the University of California is not located in Westville, there has consequently been no prolonged history of animosity, and the university is not a major political force in the community. It was realized, of course, that I was a "college professor," but nobody seemed to care which college it was.

Along with being a "real" city, Westville is reputed to have an exemplary machinery for administering criminal justice. The prosecutor's office, the police department, and the office of the public defender are generally of as high quality in facilities, pay, and national renown as those of any middle-sized city in the United States. The city jail is bright and clean. Courtrooms are modern, well-lighted, and air-conditioned. The police are housed in a modern building and are technically well equipped. In testimony before an advisory committee to the U. S. Commission on Civil Rights (1963), the head of the Westville NAACP, while stressing the continuing need for communication and understanding, said, in response to a question concerning rapport between the police department and the NAACP:

> I have the feeling that as a total police organization the Westville Police Department is head and shoulders above any other law enforcement agency in the region. This is not to say that they are perfect, or that they have even begun to approach their responsibility. They have within their Department many individuals who make . . . oppressive actions. But, at the same time, they are also enjoying an influx of new personnel . . . and the people within the Department are taking more of a professional attitude toward their responsibilities to the community.

The salutary reputation of Westville's criminal justice machinery extends throughout the United States and even abroad. During my visit to Eastville, several high-ranking officials of the police department there expressed strong interest in coming to Westville to learn its operating procedures. Consequently, since Westville is generally regarded as a model of efficiency and modernity, its administration of criminal law cannot be claimed to be representative of the United States as a whole. On the contrary, it would be more accurate to consider it as an example of the top stratum of American criminal justice administration.

THE ACTION PERSPECTIVE

To conduct research with any group whose norms and values may not conform with those of the observer, the latter must empathize with the situation of his subjects. In trying to do this, I took William F. Whyte as a model participant-observer.[1] At the time Whyte made his observations, his subjects would probably have been considered young hoodlums by the tenets of conventional morality. Whyte did not, however, *judge* his subjects, but rather observed them with the intent of understanding the bases of their behavior. As a result, the corner boy emerges as a complex human being. He appears as an actor responding to a culture in conflict at points with a conventional morality that permits certain kinds of people to be exploited and others to be treated with considerable tenderness and concern. Most importantly, the corner boy is seen by Whyte as a human being trying to work out a life for himself.

---

[1] See the Appendix, "Evolution of Street Corner Society," in William Foote Whyte, *Street Corner Society: The Social Structure of An Italian Slum* (Chicago:

Similarly, in studying the policeman the sociologist must not, if he is to develop scientifically useful hypotheses, merely emphasize the extent police behavior varies from legal ideals. Those who are of a liberal or left persuasion may find it easier to identify with the "underdog" than with constituted authority, especially the police. A participant-observer, however, must genuinely be prepared to see the world through the eyes of his subjects. Such a position does not undermine scientific objectivity, unless science is limited to the narrowest sort of positivism. On the contrary, scientific knowledge should be enhanced, provided the participant-observer is neither captured nor repelled by his subjects.

In making this investigation, therefore, I took a cue partly from sociology and partly from legal philosophy. From sociology I borrowed the "action perspective," the idea that the investigator can best perceive the meaning of events through the eyes of the participant. From legal philosophy, I borrowed Fuller's conception of law as an enterprise, a way, not always certain, of trying to reach ideals. Law is not only problematic in that its contours are vague; it is also problem-solving. It is a mechanism through which human beings, officials of varying morality and competence, struggle to solve exceedingly grave and complex problems and to balance such fundamental and conflicting ideals as order and liberty.

The legal world of rules may deceive the fresh observer into an erroneous tendency to overestimate their regularity and to consider the operation of laws as "mechanical" since they are written and need not be inferred. In reality, however, the rules of a legal order are always problematic, since the occasions for and outcomes of invoking them

University of Chicago Press, 1964), pp. 279–358; cf. also Howard S. Becker, "Problems of Inference and Proof in Participant Observation," *American Sociological Review*, 23 (1958), 652–660; and Morris S. Schwartz and Charlotte Green Schwartz, "Problems in Participant Observation," *American Journal of Sociology*, 60 (1955), 343–353.

vary. The legal philosopher H. L. A. Hart has eloquently expressed an underlying assumption used in observing the actors in this study when he speaks, in a somewhat different context, of "the open texture of law." Hart says:

> It is a feature of the human predicament . . . that we labour under two connected handicaps whenever we seek to regulate, unambiguously and in advance, some sphere of conduct by means of general standards to be used without further official direction on particular occasions. The first handicap is our relative ignorance of fact: the second is our relative indeterminancy of aim. If the world in which we live were characterized only by a finite number of features, and these together with all the modes in which they could combine were known to us, then provision could be made in advance for every possibility. We could make rules, the application of which to particular cases never called for a further choice. Everything could be known, and for everything, since it could be known, something could be done and specified in advance by rule. This would be a world fit for "mechanical" jurisprudence.

"Plainly," adds Hart, "this world is not our world." [2]

Once having observed men in a legal setting, an opposing tendency may arise. When it becomes clear that laws are not nearly so certain as was assumed, and that organizational and situational requirements often affect the actor's interpretation of laws, the sociologist may tend to forget about the rules and to interpret behavior almost purely as a response to other situational factors. The proper interpretative path here is tricky but always exciting to traverse: to see rules as a context for the behavior of legal men. As Alexander M. Bickel has stated, we are governed not by laws but by *men* who make laws:

> A government of laws simply means that no capricious commands are issued; that rules which are laid down to govern conduct are of general application; that no one man or one institution has power unchecked by anyone else; and that the lawmakers sometimes surrender immediately desired objectives for the sake of conforming to a superior norm of principle, which is itself, of course, also man-made.[3]

A study of law in action, whether of judges, lawyers, or policemen, is a study of men interpreting and thereby transforming principles and associated rules within legal institutions. Above all, it is a study of men at work. The action perspective is intended to reveal the meaning of the work to the men performing it.

---

[2] H. L. A. Hart, *The Concept of Law* (Oxford: Oxford University Press, 1961), p. 125.

[3] *The New Republic*, **151** (October 10, 1964), 9.

THE ADVANTAGE OF AN OVERVIEW

Although this book is specifically about police, it is also about other officials, the defense attorney, the prosecutor, the judge, the probation officer, because they too are woven into the system of justice without trial. A methodological conclusion of the present work is that the sociologist gains a more adequate understanding of the police by examining the work of these other officials in the system. For example, to estimate the extent of illegal police activities of various kinds, police reports alone cannot be relied on. All police have enemies, and the natural enemies of the policeman are the defense attorney and his client. Indeed, an important reason for studying a criminal law *community* is that each segment tends to be more critical of the others than of itself. From a methodological point of view it does not matter whether such criticism is justified. What matters is that natural antagonisms provide leads for investigation; more important, the absence of criticism suggests the investigator may take seriously assertions of innocence of the segment being studied. Thus, if police say there is no bookmaking or numbers racket in a city, and this assertion is confirmed by conversations over a period of several years with defense attorneys, clients, police, and court reporters, as well as random citizens in the community (who are critical of the police in other respects), it may be confidently inferred that a large-scale gambling racket does not exist. This is the basis for the conclusion that Westville does not presently have organized graft.[4]

It is, incidentally, not very difficult (as Lincoln Steffens demonstrated) to find out whether police take graft in a city. If the researcher approaches police as someone who, in effect, has worked in another police department, he finds that they, like defendants who "cop out" thinking interrogators know more than they often successfully pretend to know, will reveal themselves. There is, however, *little* revelation actually involved. In Eastville, for example, graft is routine. The police realize that a social researcher or reporter (as contrasted with a *legal* investigator who wants names and dates and places) can easily learn about the general pattern. Awareness of petty graft and corruption is necessarily widespread. If there is bookmaking, at least everybody (and his brother) who bets on horse races knows the bookmaker exists. Similarly, if numbers are being sold, knowledge of this must also be widely distributed.

---

[4] The reasons for absence are complex and deserve a study in themselves. It may be noted in passing, however, that the proximity of Nevada to California draws off much gambling money. Westvillers who might otherwise bet with a bookmaker are attracted to spend their weekends at the neon glitterings of Reno and Las Vagas.

So long as the policeman trusts the reseacher to protect his anonymity, the policeman has little to fear.

Even more important than the tactical advantage of being able to verify claims of participants is the overview an observer gains by having a form of access to the system unavailable to each of the individual participants. Under these conditions, the investigator may participate in events closed to interested actors whom he has been observing in the course of their work. For example, the "dealing out" of cases against prostitutes may take place in the chambers of a judge from which the police are excluded because their conduct is at issue in the bargaining. The researcher's presence in the judge's chambers may be of great advantage for observing how the work of the police comes to be evaluated by other actors in the system of justice without trial.

A broad overview, with access to different parties, cannot but help to aid an observer in making sociological interpretations of informal processes as well as those legally prescribed within the system. A participant-observer constantly finds himself involved in the business of interpreting the meaning of the behavior of the actors whom he is observing, and, as a corollary, the meaning the events they are engaged in present to them. Indeed, success as a sociological investigator may be measured primarily according to how well this is done, since the meaning of legal action is hardly obvious.

For example, the preliminary hearing before a magistrate is ordinarily considered as that point in the criminal procedural process at which the State must show to the satisfaction of a magistrate that there is sufficient evidence against the accused to warrant his standing trial for the crime with which he is being charged. Prosecutors and defense attorneys who were observed and interviewed during the course of preliminary hearings all felt that in *most* cases the real purpose of having a preliminary hearing, which the defendant may waive, was *not* to convince a judge there was sufficient evidence against the accused to warrant that he stand trial. That question has usually been settled to the satisfaction of both parties by the time the preliminary hearing is conducted. The prosecutor tends to find the preliminary hearing useful as an opportunity to observe how well the key State's witnesses hold up under cross-examination. On the other side, the defense attorney frequently uses the preliminary hearing as a "fishing expedition." When a defense attorney uses this expression, he suggests that in part the preliminary hearing allows him to seek actual flaws in the testimony of the State's witnesses. It also enables him to search for any statement by the witness under oath, in hope that it will later be contradicted by the same witness during the trial, thereby impeaching the validity of the witness's trial testimony.

In addition, the preliminary hearing may be used as a device by which the defense attorney controls a client reluctant to enter a guilty plea. Since a good part of the prosecutor's case may be revealed at the preliminary hearing, this step in the legal process, ostensibly designed to enable a magistrate to reach a decision, also serves as a mechanism through which the State can communicate its case to the accused. An actual demonstration of strength compels the defendant to face the reality of his legal situation, which may have been misread and misinterpreted to him by "jailhouse lawyers," and aids the defense attorney in controlling his client's agreement to plead guilty. Thus, to place this event in an "intelligible and more inclusive context of meaning," the sociological interpreter is advantaged by observing both sides of the adversarial structure.

DEVELOPMENT OF DATA COLLECTION

The study began early in the summer of 1962. I saw my task at that time as familiarizing myself with the city, the county, and especially the organization of its criminal courts community. Appendix C deals with this subject in detail. Let it be noted here as an introduction that there are two superior courts at the county level, where felony cases are tried, and that 61 per cent of all felony arrests in 1962 arose out of Westville. The public defender of La Loma County handled about 61 per cent of the 1,893 felony cases in the county. There are also 7 municipal courts which hold preliminary hearings and try misdemeanors. The "hard core" of the criminal courts community in La Loma County consists of about 40 district attorneys of varying rank, some 15 attorneys in the public defender's office, approximately 20 to 25 private defense attorneys of whom no more than 6 make a living primarily from the practice of criminal law, and 15 judges. Thus, in a county of almost one million population, the criminal courts community is very much a face-to-face one. Everybody soon knows everybody else, and most can draw upon a history of mutual relationships to estimate the merit of opinions and contentions.

It is not only size but the need for repeated interaction which makes this a face-to-face community.[5] Most of the time, a practicing attorney does not do the same kind of work that a scholar does. He has a case on his hands, and his job is to dispose of it as expeditiously as possible. Since most of his cases are settled without trial, the quality of his "contacts" will to a marked degree determine his behavior and his success as an attorney. For a practicing criminal lawyer, "knowing the ropes" often reduces to knowing people; and since interactions are necessarily frequent, he comes to know many people in a relatively short time. For instance, I found that after several months of participating in the Westville criminal courts community, I knew more people there than in my own department at the University of California; and certainly I knew people in the criminal courts community better than colleagues in such related fields as political science, law, history, anthropology, economics, and psychology. This statement may say something about the impersonality of the University of California, but I believe it reveals more about the difference in occupations. Scholarship is essentially a lonely calling. When scholars talk with one another, it is usually a social matter or results from assignment to a committee. The work of criminal lawyering, by contrast, demands repeated working contacts with others in the criminal courts community. To carry the academic analogy through, the work of criminal lawyering is like repeated assignment to an *ad hoc* committee judging the merits of an issue. The difference is that in the criminal law community, one issue is predictably repetitive: the guilt of a citizen accused of crime.

---

[5] Indeed, during the course of this study, a "Criminal Courts Association" was formed, on grounds that all parties concerned with the administration of criminal law should have a common meeting group. Although the charter members are lawyers, their feeling is that they tend to have more in common with others in the criminal courts community, such as probation officers, than with, for example, corporation lawyers. Probation officers are not permitted to become members but are asked to attend specific meetings.

The study commenced with observation of the work of the public defender in the fall of 1962. Phases of his work which were observed included: initial interviewing of the defendant; his classification as a particular category of offender; processing of the case through several stages of criminal procedure, such as arraignment, preliminary hearing, and so forth; re-evaluation of the relative strength of the defendant's case and the prosecution's on the basis of information related to these stages; communication to the defendant as to what his case is "worth" and advice as to what charge he should plead to; the defendant's reaction; the working out of a "deal" with the prosecutor; and relations between public defender and defendant, probation officer, judge, and prosecutor after the plea has been entered. Approximately two hundred hours were spent in the office of the public defender.

After several months of such observations, I felt that I had not adequately experienced the law enforcement side of the criminal courts system. It became evident that interactions between the defense attorney and the accused, as individuals, were necessarily more infrequent than those between defense attorney and prosecutor. Thus, an appropriate vision of the system of processing of criminal cases places the prosecutor and the defense attorney at the top as spokesmen and interpreters for the real adversaries who are, on the one hand, the complainant, and on the other, the accused. Given this conception, it seemed necessary to see the system of criminal law processing from the law enforcement side.

By this time, I was fairly well known to several of the deputy district attorneys who had met me while I was looking over the shoulders of men on the public defender's staff. I suggested to the head of the public defender's office that I would like to see how "the other half" lived. Through his recommendation, plus an extended interview with the district attorney, I was permitted to become a participant-observer in that office. After several weeks of observation in the office of the Westville district attorney, it seemed important to know more about the work of the police. It turned out, for instance, that policemen were often complainants, as in traffic or narcotics cases, and that even when they were not, they often represented the actual complainant. Thus, if A claimed that B had stolen his watch, a detective, after interrogating B, would talk over the case with a deputy district attorney to decide with what offense B should be charged. It became apparent that the quality of the relations between policeman and suspect influenced the working relations between policeman and prosecutor. With this in mind, I asked my "contacts" in the prosecutor's office if it would be possible to arrange observation of the police carrying out their duties.

The Chief of Police was willing to entertain the idea. It is again important to emphasize that this police department regarded itself as exemplary. It had, about a decade earlier, been a tainted police department, but it was now sure of its honesty and efficiency. The Chief was known as a man who ran a taut ship. Not beloved by either his men or the community, he was generally respected as hard-working, intelligent, and honest, although perhaps a bit cold, aloof, and sensitive to direct criticism. His attitude toward the research was made up partly of cooperation and partly of the defiance of a sensitively placed public official who feels he has done a commendable job and can say, "Go ahead and look at anything you want. Not only do we have nothing to hide in this department, but we are desirous of having our story told. We don't want a whitewash, but we do want you to be objective and truthful." On these terms my observations of the Westville Police Department began.

The Chief assigned his aide, Lieutenant Doyle, to make introductions within the department. The Lieutenant was a genial man who had been on the force for almost twenty years, knew everybody, and was personally liked, as I later learned, throughout the department. We decided that the best place to begin the study was with the patrol division which, in Westville, has one-man vehicles and three ranks: supervising detective, sergeant, and patrolman.

I spent eight nights with these patrolmen, mostly on weekends, on the shift running from 7 P.M. to 3 A.M. All of this time was spent interviewing and observing, talking about the life of the policeman, and the work of the policeman. I understood my job was to gain some insight and understanding of the way the policeman views the world. I found that the most informative method was not to ask predetermined questions, but rather to question actions the policeman had just taken or failed to take, about events or objects just encountered, such as certain categories of people or places of the city.

I soon learned, however, that patrol work is minimally connected with legal processing. To be sure, some street behavior is relevant to the policeman's role as a legal actor. On the street, the policeman has the greatest potential for discretionary judgment not to invoke the criminal law, a decision of major legal consequence for those involved. Nevertheless, I thought that the typical activities of a patrolman were not those of a *law* officer, but rather those of a *peace* officer. This distinction, I later discovered, had already been made separately by Michael Banton[6] and Arthur Stinchcombe.[7] Thus, Banton states:

> A division is becoming apparent between specialist departments within police forces (detectives, traffic officers, vice and fraud squads, etc.) and the ordinary patrolmen. The former are "law officers" whose contacts with the public tend to be of a punitive or inquisitory character, whereas the patrolmen . . . are principally "peace officers" operating within the moral consensus of the community. Whereas the former have occasion to speak chiefly to offenders or to persons who can supply information about an offense, the patrolmen interact with all sorts of people and more of their contacts center upon assisting citizens than upon offenses.[8]

Such a distinction is, of course, preliminary to the notion of an "enforcement pattern" based upon specific police assignment.

With the realization that law enforcement is not to be found in its most significant and interesting forms on the streets, I again consulted with Lieutenant Doyle (who was most helpful and considerate throughout the study). I felt that I ought to begin to study detective work, especially the work of the vice squad, but I also felt that I wanted to learn more about the policeman's use of legal authority in mundane and routine matters. Lieutenant Doyle suggested observation of the traffic division, and also suggested spending a week or so with the traffic warrant police. As it turned out, this was a useful idea, because it gave me the opportunity, as is described in one of the chapters, to observe policemen under conditions where they have wide discretion.

---

[6] Michael Banton, *The Policeman in the Community* (London: Tavistock Publications, 1964).

[7] Arthur L. Stinchcombe, "Institutions of Privacy in the Determination of Police Administrative Practices," *American Journal of Sociology*, 69 (September, 1963), 150–160.

[8] Banton, *op. cit.*, p. 7.

In line with the more general perspective of law as a problematic phenomenon, I kept in mind that observations should shed light on the policeman as a legal actor. The study of how the work of the policeman affects his family life is a reasonable subject for a sociologist of the family. The policeman's family life, however, should be studied by the legal sociologist only if there is theoretical reason to suspect a strong relation between it and his work as an officer of the law. Family life would seem to affect only shadings of work behavior, not fundamental patterns. Thus, the most important consideration here was the projected level of generalization. I was not, in this study, concerned with *individual* differences among policemen, but rather with the issue of how working conditions associated with varying enforcement specialties give rise to distinctive attitudes and behaviors of police obliged to carry these out.

With such a perspective in mind, and having already had some background in the police department, I decided to attempt to study that portion of it which seemed to me central to an understanding of the police as legal men, and perhaps also the most difficult to study: the working of the vice control squad. From the viewpoint of the sociology of law, my interest in the vice control squad was motivated by several factors.

First, the vice control squad, as a matter of observation, was in closest contact with the district attorney's office. Partly this could be attributed to exceptionally good personal relations between the head of the narcotics squad and the head of the Westville prosecutor's office. It was also dependent, however, upon a feature of vice control enforcement discussed at some length in the text: that in vice control the policeman is typically the sole legal complainant.

Associated with this was an impression, gained from working in the prosecutor's office, that vice control officers seemed in closer contact than other detectives with some of the defense attorneys. Their position as sole complainant seemed to give them a special status as an interested party in the defendant's fate. It was therefore in the interest of the defense attorney to deal with them directly in negotiating a plea of guilty.

In addition, there were some notions about the relations between substantive criminal law and criminal procedure I wished to explore. The underlying idea was that the enforcement of vice laws, especially narcotics laws, created situations inviting policemen to violate search-and-seizure restrictions; and that, as a result, the enforcement of narcotics laws had the unanticipated consequence of calling judicial attention to the behavior of the policeman, thereby restricting his future area of working discretion. But I certainly had no more than a vague idea of how this process actually worked, especially of the differing conditions under which such a generalization might be more or less true. This, then, was the perspective with which investigation of the vice control squad was begun.

Observations were usually made on weekend nights, when most activity occurred. I also spent time during the day with vice control men, especially in court-related duties, but on investigative work as well. It is difficult to state precisely how much time actually was spent with the vice control squad. Some of the most illuminating observations were made observing vice-control police in interaction with the district attorney (at which time I was primarily observing the work of the district attorney) or in interaction with other detectives while the latter provided the base for observation.

I spent six weeks, however, directly observing the vice control squad. In addition, four weeks were spent with the burglary squad and two with robbery and homicide to compare the detective's work where there is typically a citizen complainant. Weeks of intensive observation were spaced over a period of fifteen months, during which time I would drop in at least one or two afternoons a week to keep up acquaintances. I also spent one month in the summer of 1963 studying the La Loma district attorney's office. This is the office to which felony defendants are bound over after a preliminary hearing in Westville. Thus, during three months as participant-observer in the local and county offices of the prosecutor, I frequently came into contact with police.

Under direct observation, detectives were cooperative. They soon gave permission to listen in to telephone calls, allowed me to join in conversations with informants, and to observe interrogations. In addition, they called me at home when an important development in a case was anticipated. Whenever we went out on a raid, I was a detective so far as any outsider could see. Although my appearance does not conform to the stereotype of the policeman, this proved to be an advantage since I could sometimes aid the police in carrying out some of their duties. For example, I could walk into a bar looking for a dangerous armed robber who was reportedly there without undergoing much danger myself, since I would not be recognized as a policeman. Similarly, I could drive a disguised truck up to a building, with a couple of policemen hidden in the rear, without the lookout recognizing me.

At the same time, I looked enough like a policeman when among a group of detectives in a raid for suspects to take me for a detective. (It twice happened that policemen from other local departments, who recognized that I was not a member of the Westville force, assumed I was a federal agent.) Even though I posed as a detective, however, I never carried a gun, although I did take pistol training on the police range. As a matter of achieving rapport with the police, I felt that such participation was required. Since I was not interested in getting standard answers to standard questions, I needed to be on the scene to observe their behavior and attitudes expressed on actual assignments.

One problem that this sort of research approach raises is whether an observer's presence alters the normal behavior of the police. There is no certain control for this problem, but I believe the following assumptions are reasonable. First, the more time the observer spends with subjects, the more used to his presence they become. Second, participant-observation offers the subject less opportunity to dissimulate than he would have in answering a questionnaire, even if he were consciously telling the truth in response to standardized questions. "Arguing some matters," as Whyte put it, is "part of a social pattern . . . one could hardly participate without joining in the argument."[9] The process of "arguing," discussing, especially in the setting of the police work itself, creates an air of informality when opinions seem to be more openly expressed. Third, in many situations involving police, they are hardly free to alter behavior, as, for example, when a policeman kicks in a door on a narcotics raid.

Finally, if an observer's presence does alter police behavior, I believe it can be assumed that it does so only in one direction. I can see no reason why police would, for example, behave *more* harshly to a prisoner in the presence of an observer than in his absence. Nor can I imagine why police would attempt to deceive a prisoner in an interrogation to a greater degree than customary. Thus, a conservative interpretation of the materials that follow would hold that these are based upon observations of a top police department behaving at its best. However, I personally believe that while I was not exposed to the "worst," whatever that may mean, most of what I saw was necessarily typical of the ordinary behavior of patrolmen and detectives, *necessarily,* because over a long period of time, organizational controls are far more pertinent to policemen than the vague presence of an observer whom they have come to know, and who frequently exercises "drop-in" privileges. If a sociologist rides with police for a day or two he may be given what they call the

[9] Whyte, *op. cit.*, p. 302.

"whitewash tour." As he becomes part of the scene, however, he comes to be seen less as an agent of control than as an accomplice.

THE SPECIFICATION OF THE PROBLEM

In the Hunting and Neuwirth study, the researchers were not so much interested in portraying a picture of the entire litigant population in personal injury cases resulting from automobile accidents as they were in understanding the relationship between the severity of injury sustained in an automobile accident and the propensity of the injured party to sue. "We decided," say the authors, "to deal with the relatively minor type of injury because it seemed the kind of injury which confronts the injured person with the decision of whether or not to make a claim . . ." The group of respondents (persons listed on official accident reports) were chosen not because of their representative quality but because of the likelihood that their experience (being faced with the dilemma of whether to sue) would offer insights into litiguous behavior.

In addition to the fact that the Hunting-Neuwirth reading provides us with an example of the use of official records as a means of data collection, there is one other aspect of the selection which is worth noting--the authors use the device of random sampling in order to obtain an understanding of the universe with which they are dealing.

We will examine the issue of sampling in Chapter IV. At this point, we would merely like to indicate that Hunting and Neuwirth were faced with the problem of having to ascertain information about large masses of people. Short of interviewing all these people, a random sample is the best guarantee that the information gained describes the entire group with accuracy.

From Roger B. Hunting and Gloria S. Neuwirth, Who Sues in New York City? (New York: Columbia University Press, 1962) pp. 1-7. Reprinted by permission.

Roger B. Hunting and Gloria S. Neuwirth,
Who Sues in New York City? (New York: Columbia
University Press, 1962), pp. 1-7.

The Project's goal was to discover the makeup of that part of the population of New York City which is the source of personal injury claims and lawsuits. To reach that goal, it was apparent that long and detailed interviews with injured people were necessary. There seemed no other effective way of obtaining information in the areas we wished covered, which included the following: the circumstances of the accident which caused the injury; the extent of financial loss incurred; the liability situation; the influences of others upon the injured man; previous experiences with accidents and injuries as well as lawyers and courts; attitudes toward people who make claims for their injuries, toward insurance companies, and lawyers, as well as general sophistication or degree of awareness of the system of recovery for injuries; and, finally, education, income, age, and other relevant aspects of background. It was obvious that accurate and complete answers to questions in these fields of inquiry could best be obtained through personal interviews with a scientifically selected sample of accident victims.

The information obtained from such a sample would then give us a picture of the nature of the "litigating public," and would provide a firm basis for recommendations affecting the flow of litigation into and through the courts. For example, if the research had revealed (which it did not) that drivers over the age of sixty are much more prone to sue than others, the adoption of laws restricting licenses to those under that age might be expected to reduce the flow of cases into the courts by removing a more than generally litigious group from the total of potential litigants.

The approach the research took was to obtain information about those persons who, in effect, are faced with a choice as to whether or not to make a claim or obtain a lawyer. But although a goal of this research is the reduction of the volume of cases which burden our courts, steps toward that goal must be consistent with the overriding necessity that substantial justice be done for all the citizens. Fundamentally, we are committed to the proposition that every person has the right to bring his controversy to a legal tribunal to have it decided in accordance with law. Therefore, any exclusion of controversies from easy access to a judicial body is, to some extent, a denial of justice to those who are entitled to it. It is because this is true that justice delayed is, in some cases, justice denied. It is also because this is true that we cannot be arbitrary in our efforts to reduce the volume of litigation—the right to litigate must be preserved.

Nevertheless, some restrictions on that right have been established from time to time. The most notable of these is the entire workmen's compensation system, which removed from court consideration practically all cases in which workers, injured at work, made claims against their employers for compensation for their injuries. Before that system was instituted, an injured worker was relegated to an action against his employer under the laws of tort, based on a claim that the employer had neglected his duty to provide the employee with a safe place to work or had otherwise neglected to protect his health and safety. Workmen's compensation laws now assign this area of controversy to boards, rather than the courts, and to this extent the right to litigate has been eliminated.

Generally speaking, however, thoughtful persons are interested in maintaining the right of those who wish to use court processes to do so, and the reduction of the influx of cases, if accomplished, must be by a method which limits the reduction to that portion of cases in which resort to the courts is not reasonable or justified. Obviously, a person who has been severely injured or who has lost a limb or been permanently disabled is completely justified in retaining a lawyer to submit to a court the questions of another's liability for the loss, and assessment of the damage against that other if liable. It is clear that our system must, to some extent, lean over backward to provide a forum for such cases. On the other hand, it is equally clear that if a drunken driver runs into a cement abutment and, providentially, sustains a bruised knee and nothing more, he is not reasonable or justified in demanding a court adjudication of his claim against the city that the abutment had been negligently built in a spot through which he was driving.

Attention, therefore, has been devoted to an examination of the nature and actions of people whose accidents and injuries are sufficiently slight that one person might say, "I'll forget it," while another says, "I'll sue." It was, of course, necessary to deal with a group which could be ascertained without reference to claim or suit so that we would not be limited to those who had already committed themselves to make a claim or institute suit, while missing those (we were confident of their existence) who would not, in spite of being injured, make claim or institute suit.

The only available source of a large pool of injured persons who might or might not retain lawyers and make claims was that great mass of persons who are injured each year in automobile accidents. The law requires that their names be reported to the State Bureau of Motor Vehicles, and compliance with the law provides the names of those injured, without reference to whether or not they plan to seek compensation for those injuries. It is also of great value to deal with automobile accident victims since, in addition to being the only group available for whom such records exist, they do, especially outside New York County, provide the source of the great majority of all the litigation which occupies the courts of this city and the country at large.

Thus, the decision was made to direct the study to persons who were injured in motor vehicle accidents, this group providing the greatest influx of suits resulting in calendar congestion. The total number of accidents in New York State in 1957 (a typical year) in which there was a personal injury was 162,523, and as a result of these accidents, a total of 257,934 persons were injured. Because delay in New York City courts was most troublesome, research was limited to accidents occurring in the City. This just about halved the number of accidents (in 1957) to 85,149, and the number of injured persons to 134,891.

The New York State Bureau of Motor Vehicles divides injuries into eight categories of varying degrees of seriousness, the mildest of which is recorded as "slight shock, contusions." In the words of the injured person this may be "shaken up," "bruises," or "injuries not yet known." We decided to deal with this relatively minor type of injury because it seemed the kind of injury which confronts the injured person with the decision of whether or not to make a claim, and because it may be dismissed as nothing by those with any inclination to do so. In addition, since 50 percent or 68,155 of the 134,891 injuries reported in New York City in 1957 fell within this mild category, about one half of the potential claimants and claims would be represented.

We selected the year 1957 for the sampling, since it was close enough to be within the range of memory for the persons injured, yet distant enough so that the great majority of claims brought would have been disposed of by the time the interviews were held.

A detailed description of the method by which the sample was drawn and other information on the subject appears as Appendix A. At this juncture, however, we think it worthwhile to explain briefly our sampling methods. In general terms, it was a random sample of 321 injured persons, drawn from the total injured in the group we were concerned with, whose injuries were incurred during the months of February, March, April, and May, 1957. The final sample, then, consisted of 321 persons injured in New York City in 1957 with slight shock and contusion. This is less than 1.5 percent of the total so injured in New York City in the four-month period.

Nevertheless, the technique of random sampling has been sufficiently developed and tested that it can be said with confidence that the information gained from the examination of such a scientific sample describes the entire group with some accuracy. For example, for present purposes, if one sought information about 1,000 people, a sample of 100 would be sufficient and would probably be as accurate as a sample of 200. In fact, a sample of 100 in which everyone in the sample was questioned would be more accurate than a sample of 400 of which only one half were questioned, even though that group of 200 is twice as large as the first group. The reason is that when persons are interviewed, unless strenuous efforts are made to interview the reluctant as well as the willing, a bias will creep in, simply because the consenting interviewees usually have characteristics different from those of the reluctant ones who are not interviewed. Thus, in the example given above, the 200 interviewed from the sample of 400 are likely to have characteristics which differ from those of the 200 who were not interviewed, so that the results when tabulated will be weighted in favor of the characteristics of the willing. On the other hand, if the sample of 100 were all interviewed, the characteristics of the willing would be balanced by the characteristics of the unwilling, so that a more accurate picture of the entire group (1,000) would be obtained.

Therefore, although 321 injured people is but a small portion of the total "universe" from which it was drawn, the random method of its selection substantially assures that it will accurately reflect the whole group from which it came. This was, as noted, over one half of the people injured in New York City and thus over one half of the potential claimants and litigants.

Of the 321 persons in the sample, 165 were actually interviewed and 164 were coded and analyzed. That is 51 percent of the total. Of those not interviewed, 72 were totally unavailable, having moved and left no address, being unknown or having given a fictitious address, or having attorneys who refused to permit an interview. Thus, of the 249 who were in fact available to the interviewers, 66 percent were interviewed. Interviews were completed by September 15, 1959 and our discussion throughout reflects the situation at that date.

The analysis of the interviews gave us what might be described as a claimant's-eye view of the entire process of compensation for accident victims from the initial accident to a final settlement or trial resulting in payment. The information we sought was really the answer to one rather simple question: What distinguishes the person who makes a claim from the one who does not, and if he makes the claim, what distinguishes the person who retains a lawyer from the one who does not?

Specifically, we obtained information as to these distinguishing features in the areas of amount of loss; evaluation of liability; influence of others; previous experience with accidents, lawyers, and courts; attitude toward courts, lawyers, insurance companies, etc.; and the socio-economic background of the injured person.

OFFICIAL RECORDS

The following selection from the Wolfgang deposition provides an example of research design not based on participant observation or surveys. In addressing the question of race discrimination in sentencing for rape, Wolfgang tells us that he "decided to start with the only reliable and objectively ascertained fact that was available . . . the fact of conviction for the crime of rape." Official records are a convenient source of information, and in some instances they are the only practical source of data available. They are often of limited value, however.

In planning his research design, Professor Wolfgang was aware of the limitation imposed by relying on official records. "Using this point of departure [official records] meant," he points out, "that we would not be able to say, if at the end of the process a disproportionate number of Negroes were found sentenced to death, whether the disproportion was in part the result of rapes by Negroes being disproportionately frequently reported to the police, or of Negroes being disproportionately frequently arrested for rape or charges with rape, etc. . ." In short, he was not looking for a comprehensive picture of the effects of race on the criminal process but was interested in one very limited question--namely, the impact that race has in determining sentencing among all persons convicted of the crime of rape.

None of this is to say, of course, that we are not in favor of the use of official records in planning a research design. Rather, as we suggested in our introduction to this chapter, the best research design is one which tries to obtain information in a variety of ways, both to serve as a cross-check and to increase the scope of our information.

From Marvin E. Wolfgang, deposition in case of Alabama v. Wheeler Billingsley, State of Alabama for Etowah County (Cases 1159, 1160, 1161).

THE STATE OF ALABAMA

ETOWAH COUNTY

THE STATE OF ALABAMA,

Plaintiff

vs.

WHEELER BILLINGSLEY, JR.
JAMES LIDDELL and ROBERT BUTLER,

Defendants

CASE NOS. 1159, 1160, 1161

Deposition of Marvin E. Wolfgang witness sworn and examined on behalf of Wheeler Billingsley, Jr., James Liddell and Robert Butler by virtue of the annexed commission issued out of the Circuit Court for Etowah County and State of Alabama, in a case therein pending, wherein the State of Alabama is Plaintiff and Wheeler Billingsley, Jr., and James Liddell and Robert Butler are Defendants. The said Marvin E. Wolfgang first being duly sworn to speak the truth, the whole truth and nothing but the truth, doth depose as follows:

To the nineteenth interrogatory, he saith:

My colleague at the University of Pennsylvania, Professor Anthony Amsterdam of the Law School, asked me if I would be interested in participating in a research project to study the effects, if any, of racial factors in the imposition of the death penalty upon defendants who have been convicted of the crime of rape in Southern states. He stated his belief that this was an appropriate topic for research and analysis in light of the apparent fact -- as suggested by the National Prisoner Statistics on Executions published by the Federal Bureau of Prisons -- that a far greater number of Negro defendants receive the death penalty for rape in the South than do white defendants. The Bureau's figures show, specifically, that of the 446 persons executed for the crime of rape in the United States between 1930 and 1962, 399 were Negro, 45 were white and 2 were of other races. I am attaching as Exhibit 1-A to this deposition a xerox copy of tables in the Bureau's publication, National Prisoner Statistics, Executions 1962, in which these figures appear. The 9-to-1 Negro-white ratio in executions for rape was particularly interesting to us inasmuch as the Negro-white ratio in executions for murder during the same period was about 1-to-1 and because rape is punishable capitally almost exclusively in the Southern and border States. I was asked to help construct a study design that would provide the basis for determining whether this apparent disproportion in the sentencing of Negroes to death for rape was real, and whether it resulted from differential sentencing practices based on race or was due to other causes.

To the twentieth interrogatory, he saith:
I was informed by Professor Amsterdam that the sponsor of the project was the NAACP Legal Defense and Educational Fund, Inc., and that if the proposed study was in fact carried out, the results of the study might be used in litigation in cases in which lawyers working for the NAACP Legal Defense and Educational Fund, Inc. were representing Negro defendants who have been convicted of rape and sentenced to death.

To the twenty-first interrogatory, he saith:
I was asked initially if I would attend a conference to discuss the appropriate and practicable scope of a possible research project to study the influence of racial factors in sentencing for rape, and the means of setting up such a research project in a fashion that would assure scientifically reliable and valid results. I was asked to serve as the technical consultant in formulating the comprehensive research design for the project. I was asked to give instructions to researchers who would collect empirical or field data pursuant to the research design. I was also requested to do a statistical analysis of the data collected as a consequence of the field research. I acceded to these requests and participated in the conference, the design of the research study, and subsequent phases of the study including analysis of data.

To the twenty-second interrogatory, he saith:
The study was to be concerned principally with capital sentencing for rape. I was told that only eighteen American jurisdictions now allow the imposition of the death sentence for rape; the federal jurisdiction, District of Columbia, and the States of Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, Missouri, Nevada, North Carolina, Oklahoma, South Carolina, Tennessee, Texas and Virginia.

The National Prisoner Statistics of the Federal Bureau of Prisons disclosed that substantial numbers of persons had been executed for the crime of rape during the past 30 years or so in only twelve of these States: Alabama, Arkansas, Florida, Georgia, Louisiana, Maryland, Mississippi, North Carolina, South Carolina, Tennessee, Texas and Virginia. The State on this list with the fewest rape executions between 1930 and 1962, Louisiana, had 17; no state not on the list had more than 10. This appears from the National Prisoner Statistics table #4 attached as Exhibit 1-A. It was decided that the study should be profitably confined to these twelve States.

I suggested that, to form a basis for conclusions about the effect of racial factors on capital sentencing for rape in each of these States, it would be necessary to gather data about a sizeable or substantial number of rape cases in each State. In order that satisfactory statistical analysis be performed, it is necessary to have a considerable number of instances of any phenomenon that is being studied. Here, the phenomenon studied was the type of sentences imposed upon convictions for rape. That meant we had to gather information, in each State, about a considerable number of cases of conviction for rape.

I should say that we decided to start with rape convictions in studying the effect of race on capital sentencing for rape because there was no other practicable place to start. In seeking to explore the meaning of the apparent racial disparity in capital sentencing for rape revealed by the National Prisoner Statistics -- that is, the phenomenon that nine times as many Negroes as whites were executed for rape in the United States between 1930 and 1962 -- one might well be led to ask the question whether Negroes commit rape more frequently than whites. But it was not possible, as a practical matter, to answer this question, because it was not possible to determine the total number of rapes committed or the identity and race of the persons who committed them. Many rapes -- an unknown number -- are not reported to the police, and of those that are reported, many are not cleared by arrest. Even when a reported rape is cleared by arrest, often no records of any accuracy or completeness are kept. So no reliable information would be available on the total number of rapes committed or who committed them.

Similarly, it was not practicable to answer questions about the number of Negroes and whites charged with rape. In numerous cases where a charge is made, it is dropped or the person charged is acquitted. If the charge is dropped without a trial, it is usually very difficult to get much information about the circumstances of a case. And, in any event, it is almost impossible to determine the significance that should be given to a case in which the disposition is a dropping of the charges or a verdict of acquittal. Should we find that an equal number of Negroes and whites were charged with rape, but that nine times as many Negroes were eventually sentenced to death for the crime, this might mean only that nine times as many Negroes as whites were guilty of the crime, and that, in effect, whites were much more frequently falsely charged with rape than Negroes. To determine the validity of this latter speculation, it would be necessary to find some means of deciding whether each defendant who was charged with rape was in fact guilty or not guilty -- that is, of deciding whether the prosecutor or the judge or jury had rightly determined guilt or innocence. Even if we knew the facts of all the cases -- which, as I have said, we would not likely know in many cases where prosecutions were dropped in their early states -- we would have no objective and reliable means for resolving conflicting evidence, for example, or for drawing the kind of inferences which prosecutors, judges and juries draw from evidence in deciding guilt or innocence in particular cases.

For these reasons, we decided to start with the only reliable and objectively ascertainable fact that was available to us; the fact of conviction for the crime of rape. We decided to study the cases of persons convicted of rape, in order to learn whether, among the total number of such persons, race was a factor determining who would be sentenced to death and who would be sentenced to something less than death. Using this point of departure meant, of course, that we had to ignore the effect of racial factors upon the criminal process prior to the conviction; we would not be able to say, if at the end of the process a disproportionate number of Negroes were found sentenced to death, whether the disproportion was in part the result of rapes by Negroes being disproportionately frequently reported to the police, or of Negroes being disproportionately frequently arrested for rape or charged with rape, or of prosecutors disproportionately frequently dropping rape charges against whites, or of juries disproportionately frequently convicting Negroes. One might like to know, of course, in order to obtain a comprehensive picture of the role which race plays in the administration of criminal justice in rape cases, whether racial factors affect all of these stages of the process. But it was not practicable to study these stages, and we therefore determined to limit our study to the narrower but very important question of the influence of race at the sentencing stage: that is, what impact race has in determining sentence among all persons convicted of the crime of rape.

In doing so, of course, we automatically negatived the speculative possibility that, if Negroes were disproportionately sentenced to death for rape, that pattern could be accounted for by a disproportionate frequency in the commission of rapes by Negroes, or in the conviction for rape of Negroes. Starting with a class of cases in all of which the defendant was convicted of rape, we could investigate whether among convicted defendants, Negroes were disproportionately frequently sentenced to death and, if so, whether the disproportion seemed to be explained by non-racial factors.

As I said, then, we had to have in each State to be studied a considerable number of cases of convictions for rape. We could not know in advance how many there would be in any given state within any given period. We assumed that we could find a substantial number by looking at every case of conviction for rape in any state over a relatively long period of time, say ten years, or by looking at every case of conviction for rape in selected counties within that state for a somewhat longer period of time. We decided on the latter course. We had limited resources available for the study and wanted to utilize them most efficiently. A given number of field researchers can gather more information about more cases in a restricted number of counties for a longer period than in a larger number of counties for a shorter period, because moving from county to county involves considerable time lost in travel, locating and becoming familiar with new records, and other transition problems. Studying selected counties in a state during a longer period, rather than all counties in the state during a shorter period had the added advantage of allowing us to examine the development over time of racial influences on sentencing. And, if the counties to be studied were selected by a proper sampling procedure, so that the sample counties were representative of the state as a whole, the conclusions drawn from the data collected in those counties would be valid for the entire state.

We decided, therefore, to study every case of conviction for rape during a twenty-year period -- January 1, 1945, until the summer of 1965 when the study would be undertaken -- in about 250 counties comprising a sample of the counties in each of twelve states, and comprehending more than 50% of the total population of those twelve states. Actually, as it turned out, we did not have time to do Maryland, and so gathered data from about 230 counties in eleven states.

To the twenty-third interrogatory, he saith:

In order to obtain a reliable determination of the effect of racial factors on sentencing for rape, it was necessary first to construct the questions which were the subject of inquiry -- whether, for example, Negroes convicted of rape in the South were disproportionately frequently sentenced to death in comparison with whites, and, if so, whether the disproportion was attributable solely to race or rather explicable in terms of other factors -- into a scientifically testable hypothesis or hypotheses. By that I mean it was necessary to construct hypotheses in terms such that they could be tested by observation, by the collection of factual data, which, when analyzed statistically, would yield results confirming or rejecting the hypotheses. Most empirical scientific research ordinarily begins with this sort of hypotheses. Their function, in the first instance, is to direct the factual investigation by identifying what data need to be observed or collected. The hypotheses, in testable form, directs attention to the specific facts which will sustain or refute them. These are the facts about which data must be obtained.

It was appropriate to translate our ultimate question concerning the effect of race on sentencing into subsidiary questions. First we asked whether there was any relation between defendants' race and sentence -- that is, whether Negro defendants convicted of rape were disproportionately frequently sentenced to death, compared with white convicted defendants; and then we asked whether Negro defendants convicted of raping white victims were disproportionately frequently sentenced to death compared to whites who raped whites, whites who raped Negroes, and Negroes who raped Negroes. If the answer to either of these questions was yes, we had to ask a second sort of question: whether any non-racial factor or factors could account for or explain the racial disproportions.

Take the first kind of question first, and--as an example--the specific question whether Negroes convicted of rape are disproportionately frequently sentenced to death compared to whites. To make this question scientifically testable, we first framed it as a null hypothesis. A null hypothesis is one which states that there is no relationship between factors or, put another way, that there is no differential distribution of the data under observation. The particular null hypothesis which was framed to test this first question, then, is that there is no significant relationship between the race of a defendant and sentence--that there is no significant difference between the proportion of Negro defendants convicted of rape who are sentenced to death and the proportion of white defendants convicted of this crime who are sentenced to death.

The study was designed, as stated above, to provide a basis for testing this hypothesis. The hypothesis is in a form such that it can be tested by collecting data which when analyzed will provide a comparison between the expected distribution of death cases--that is, the frequency of death sentences which would be imposed on Negroes and the frequency of death sentences which would be imposed on whites if the null hypothesis was valid--and the observed distribution of cases in which the death sentence actually was imposed. If the observed data show that a substantially or significantly different number of Negro and white defendants were actually sentenced to death than the numbers one would expect under the null hypothesis (the hypothesis that Negroes are not disproportionately frequently sentenced to death), then the hypothesis is rejected and one can conclude that Negroes are disproportionately frequently sentenced to death.

It was in this fashion that we framed as hypotheses all of our questions of both sorts described above--concerning the relation of race and sentence, and also concerning the relation of other, possibly explanatory variables to sentence. These hypotheses told us what data we had to collect. . . .

SABOTAGE AND THE HIRED HAND

The following article "Hired Hand Research" is, we feel, one of considerable importance in that it speaks to an all too often underplayed problem in social science research. Professor Roth says: "There is no reason to believe that a hired hand in the scientific research business will behave any differently from those in other areas of productive activity. They want to make as much money, avoid difficult, embarrassing, inconvenient, time consuming situations as well as those activities which make no sense to them." In short, what Roth is saying is that hired help employed to interview, program, analyse statistics, etc., can sabotage even the most flawless of research designs.

For the lawyer, Roth's article has particular relevance. He will not only have to contend with the hired hands Roth refers to, but more than likely will find that he has to collaborate with a social scientist in planning research design, constructing hypotheses, and drawing samples. It is conceivable, then, that the lawyer might find himself in the position of having employed a behavioral scientist, who, as Roth points out, is apt to "avoid dealing with activities which make no sense to him."

What is essential, if this hired hand syndrome is to be avoided in interdisciplinary research, is that lawyer and scientist, as did Wolfgang and Amsterdam, have not only a common interest but a common understanding of the problem which is being confronted.

From Julius A. Roth, "Hired Hand Research," American Sociologist, (August, 1966) pp. 190-196. Reprinted by permission.

Julius A. Roth, "Hired Hand Research," American Sociologist, (August, 1966), pp. 190-196.

CASE I

After it became obvious how tedious it was to write down numbers on pieces of paper which didn't even fulfill one's own sense of reality and which did not remind one of the goals of the project, we all in little ways started avoiding our work and cheating on the project. It began for example when we were supposed to be observing for hour and a half periods, an hour and a half on the ward and then an hour and a half afterwards to write up or dictate what we had observed, in terms of the category system which the project was supposed to be testing and in terms of a ward diary. We began cutting corners in time. We would arrive a little bit late and leave a little bit early. It began innocently enough, but soon boomeranged into a full cheating syndrome, where we would fake observations for some time slot which were never observed on the ward. Sam, for example, in one case, came onto the ward while I was still finishing up an assignment on a study patient and told me that he was supposed to observe for an hour and a half but that he wasn't going to stay because he couldn't stand it anymore. He said he wasn't going to tell anyone that he missed an assignment, but that he would simply write up a report on the basis of what he knew already about the ward and the patients. I was somewhat appalled by Sam's chicanery, and in this sense I was the last one to go. It was three or four weeks after this before I actually cheated in the same manner.

It was also frequent for us to miss observation periods, especially the 8 to 9:30 a.m. ones. We all had a long drive for one thing, and we were all chronic over-sleepers for another. For a while we used to make up the times we missed by coming in the next morning at the same time and submitting our reports with the previous day's date. As time went on, however, we didn't bother to make up the times we'd missed. When we were questioned by our supervisor about the missing reports, we would claim that there had been an error in scheduling and that we did not know that those time slots were supposed to be covered.

There were other ways we would cheat, sometimes inadvertently. For example, one can decide that one can't hear enough of a conversation to record it. People need to think fairly highly of themselves, and when you think that you're a cheat and a liar and that you're not doing your job for which you are receiving high wages, you are likely to find little subconscious ways of getting out of having to accuse yourself of these things. One of the ways is to not be able to hear well. We had a special category in our coding system, a question mark, which we noted by its symbol on our code sheets whenever we could not hear what was going on between two patients. As the purgatory of writing numbers on pieces of paper lengthened, more and more transcripts were passed in with question marks on them, so that even though we had probably actually heard most of the conversations between patients, we were still actually avoiding the work of transcription by deceiving ourselves into believing that we could not hear what was being said. This became a good way of saving yourself work. If you couldn't hear a conversation, it just got one mark in one column of one code sheet, and if you wrote down an elaborate conversation lasting even ten minutes, it might take you up to an hour to code it, one hour of putting numbers in little blocks. In the long run, all of our data became much skimpier. Conversations were incomplete; their duration was strangely diminishing to two or three minutes in length instead of the half-hour talks the patients usually had with each other. We were all defining our own cutting off points, saying to ourselves, "Well, that's enough of that conversation." According to the coding rules, however, a communication can't be considered as ended until the sequence of interaction has been completed and a certain time lapse of silence has ensued.

In order to ensure the reliability of our coding, the research design called for an "Inter-Rater Reliability Check" once every two months, in which each of the four of us would pair up with every other member of the team and be rated on our ability to code jointly the same interaction in terms of the same categories and dimensions. We learned to loathe these checks; we knew that the coding system was inadequate in terms of reliability and that our choice of categories was optional, subjective, and largely according to our own sense of what an interaction is really about, rather than according to the rigid, stylized, and preconceived design into which we were supposed to make reality fit. We also knew, however, that our principal investigators insisted on an inter-rater reliability coefficient of .70 in order for the research to proceed. When the time came for another check, we met together to discuss and make certain agreements on how to bring our coding habits into conformity for the sake of achieving reliability. In these meetings we would confess our preferences for coding certain things in certain ways and agree on certain concessions to each other for the duration of the check. Depending on what other individual I was to be paired with, for example, I had a very good idea of how I could code in order to achieve nearly the same transcriptions. We didn't end it there. After each phase of a check, each pair of us would meet again to go over our transcriptions and compare our coding, and if there were any gross discrepancies, we corrected them before sending them to the statistician for analysis. Needless to say, as soon as the reliability checks were over with, we each returned to a coding rationale which we as individuals required in order to do any coding at all—in order to maintain sanity.

CASE II

There didn't appear to be too much concern with the possibility of inconsistency among the coders. Various coders used various methods to determine the code of an open-end question. Toward the end of the coding process, expediency became the keynote, leading to gross inconsistency.

POOR ORIGINAL COPY - BEST
AVAILABLE AT TIME FILMED

The most expedient method of coding a few of the trickier questions was to simply put down a "4" (This was the middle-of-the-road response on the one question that had the most variation.). If the responses were not clear or comprehensible, the coder had two alternatives: on the one hand, he could puzzle over it and ask for other opinions or, on the other hand, he could assign it an arbitrary number or forget the response entirely.

In the beginning, many of us, when in doubt about a response, would ask the supervisor or his assistant. After a while, I noted that quite often the supervisor's opinion would differ when asked twice about the same response and he would often give two different answers in response to the same question. One way the supervisor and his assistant would determine the correct coding for an answer would be to look at the respondent's previous answers and deduce what they should have answered--thereby coding on *what they thought the respondent should have answered*, not on the basis of what he *did* answer. One example that I distinctly remember is the use of magazines regularly read as reported by the respondent being used as a basis on which to judge and code their political views. This, in my opinion, would be a factor in some of the cases, such as the reading of an extreme leftist or extreme rightist magazine, but to use magazines such as *Time* or *Reader's Digest* to form any conclusions about the type of person and his views, I feel is quite arbitrary. Furthermore, I feel questionnaires should be used to see *if* consistent patterns of views exist among respondents and it is not the coder's job to put them in if the respondents fail to!

Some of the coders expected a fixed pattern of response. I, not being sure of what responses meant in a total political profile, treated each response separately--which I feel is the correct way of coding a questionnaire. Others, as I learned through their incessant jabbering, took what they thought was a more sophisticated method of treating an interview. A few would discuss the respondent's answers as if they took one political or social standpoint as an indicator of what all the responses should be. They would laugh over an inconsistency in the respondent's replies, feeling that one answer did not fit the previous pattern of responses.

The final problem leading to gross inconsistency was the factor of time. The supervisor made it clear that the code sheets had to be in to the computation center by Saturday. This meant that on Saturday morning and early afternoon the aim of the coders was to code the questionnaires as quickly as possible, and the crucial factor was speed, even at the expense of accuracy. The underlying thought was that there were so many questionnaires coded already (that were *assumed* to be coded consistently and correctly) that the inconsistencies in the remainder would balance themselves out and be of no great importance. I found myself adapting to this way of thinking, and after spending two or three hours there on Saturday morning, I joined in the game of "let's get these damn things out already." It did indeed become a game, with the shibboleth, for one particularly vague and troublesome question, "Oh, give it a four."

CASE III

One of the questions on the interview schedule asked for five reasons why parents had put their child in an institution. I found most people can't think of five reasons. One or two--sometimes three. At first I tried pumping them for more reasons, but I never got any of them up to five. I didn't want (the director) to think I was goofing off on the probing, so I always filled in all five.

Another tough one was the item about how the child's disability affected the family relationships. We were supposed to probe. Probe what? You get so many different kinds of answers, I was never sure what was worth following up. Sometimes I did if the respondent seemed to have something to say. Otherwise I just put down a short answer and made it look as if that was all I could get out of them. Of course, (the director) *did* list a few areas he wanted covered in the probing. One of them was sex relations of the parents. Most of the time I didn't follow up on that. Once in a while I would get somebody who seemed to be able to talk freely without embarrassment. But most of the time I was afraid to ask, so I made up something to fill that space.

Then there was that wide open question at the end. It's vague. Most people don't know what to say. You've been asking them questions for about an hour already. Usually you get a very short answer. I didn't push them. I'd write up a longer answer later. It's easy to do. You have their answers to a lot of other questions to draw on. You just put parts of some of them together, dress it up a little, and add one or two bits of new information which fits in with the rest.

Any reader with research experien can probably recall one or more cas in which he observed, suspected, or participated in some form of cheating, carelessness, distortion, or cutting of co ners in the collection or processing research data. He probably thought of these instances as exceptions—an unfortunate lapse in ethical behavior or failure of research directors to maintai proper controls. I would like to put forth the thesis that such behavior o the part of hired data-collectors an processors is not abnormal or exceptional, but rather is exactly the kind of behavior we should expect from peopl with their position in a production uni

The cases I have presented do no constitute proof, of course. Even if I presented ten or twenty more, my effort could be dismissed as merely an unusually industrious effort to record professional dirty linen (or I might be accused of making them up!) and not a all representative of the many thousand of cases of hired researching carried out every year. Rather than multiply examples, I would like to take a differen tack and examine the model we hav been using in thinking about research operations and to suggest another model which I believe is more appropriate.

The ideal we hold of the researcher i that of a well-educated scholar pursuing information and ideas on problems in which he has an intrinsic interest. Frequently this ideal may be approximate when an individual scholar is working on his own problem or several colleagues are collaborating on a problem of mutual interest. Presumably such a researcher will endeavor to carry out his data-collection and processing in the most accurate and useful way that his skills and time permit.

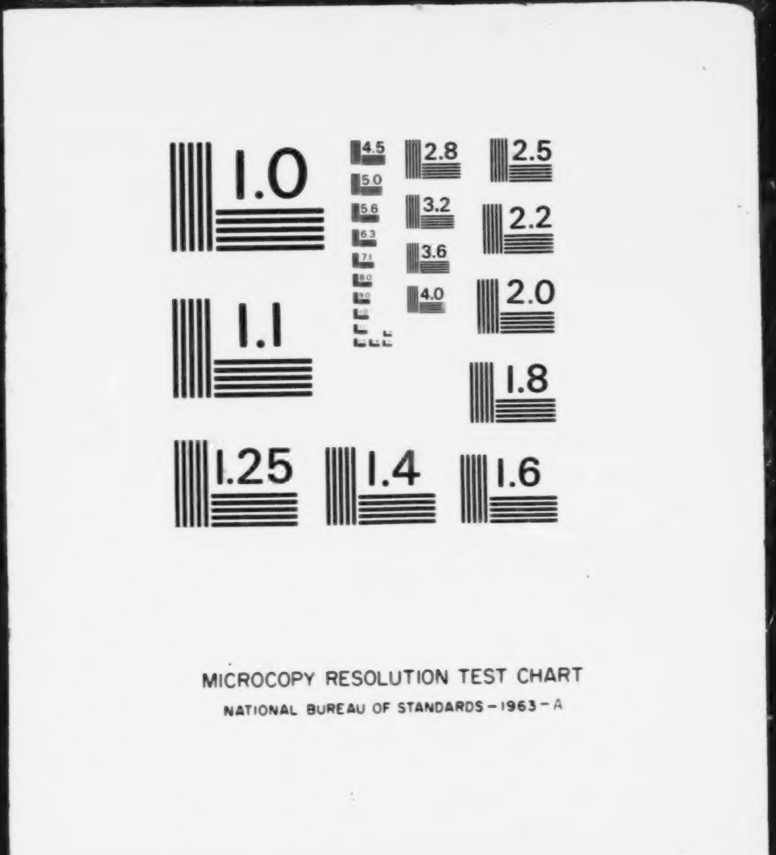
When a researcher hires others to do the collecting and processing tasks of his research plan, we often assume that these assistants fit the "dedicated scientist" ideal and will lend their efforts to the successful conduct of the over-all study by carrying out their assigned tasks to the best of their ability. As suggested by my examples, I doubt that hired assistants usually behave this way even when they are junior grade scholars themselves. It becomes more doubtful yet when they are even further removed from scholarly tradition and from the direct control of the research directors (e.g., part-time survey interviewers).

It seems to me that we can develop a more accurate expectation of the contribution of the hired research worker who is required to work according to somebody else's plan by applying

5 OF 7

ED

044322


1.0
4.5
5.0
5.6
6.3
7.1
8.0
9.0
2.8
3.2
3.6
4.0
2.5
2.2
2.0
1.1
1.8
1.25
1.4
1.6
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

another model which has been worked out in some detail by sociologists—namely, the work behavior of the hired hand in a production organization. First, let us look at one of the more thorough of these studies, Donald Roy's report on machine shop operators.[1]

Roy's workers made the job easier by loafing when the piece rate did not pay well. They were careful not to go over their informal "quotas" on piece rate jobs because the rate would be cut and their work would be harder. They faked time sheets so that their actual productive abilities would not be known to management. They cut corners on prescribed job procedures to make the work easier and/or more lucrative even though this sometimes meant that numerous products had to be scrapped. Roy's calculations show that the workers could have produced on the order of twice as much if it had been in their interest to do so.

But it is *not* in their interest to do so. The product the hired hand turns out is not in any sense his. He does not design it, make any of the decisions about producing it or about the conditions under which it will be produced, or what will be done with it after it is produced. The worker is interested in doing just enough to get by. Why should he concern himself about how well the product works or how much time it takes to make it? That is the company's problem. The company is his adversary and fair game for any trickery he can get away with. The worker's aim is to make his job as easy and congenial as the limited resources allow and to make as much money as possible without posing a threat to his fellow workers or to his own future. The company, in turn, is placed in the position of having to establish an inspection system to try to keep the worst of their products from leaving the factory (an effort often unsuccessful—the inspectors are hired hands, too) and of devising some form of supervision to limit the more extreme forms of gold-bricking and careless workmanship.

Almost all the systematic research on "restriction of output" and deviation from assigned duties has been done on factory workers, office clerks, and other low prestige work groups. This is mostly because such work is easier to observe and measure, but also because much of this research has been controlled in part by those in a position of authority who want research done only on their subordinates. However, there is evidence to indicate that work restrictions and deviations in the form of informal group definitions and expectations are probably universal in our society. They can be found among business executives and in the professions, sports, and the creative arts. They are especially likely to crop up when one is working as a hired hand, and almost all productive activities have their hired hand aspects. A professor may work hard on scholarly tasks of his own choosing and perhaps even on teaching a course which he himself has devised, but he becomes notoriously lax when he is assigned to a departmental service course which he does not like—spending little or no time on preparation, avoiding his students as much as possible, turning all the exams over to a graduate assistant, and so on.

"Restriction of production" and deviation from work instructions is no longer regarded by students of the sociology of work as a moral issue or a form of social delinquency. Rather, it is the expected behavior of workers in a production organization. The only problem for an investigator of work practices is discovering the details of cutting corners, falsifying time sheets, defining work quotas, dodging supervision, and ignoring instructions in a given work setting.

There is no reason to believe that a hired hand in the scientific research business will behave any different from those in other areas of productive activity. It is far more reasonable to assume that their behavior will be similar. They want to make as much money as they can and may pad their account or time sheet if they are paid on that basis, but this type of behavior is a minor problem so far as the present discussion is concerned. They also want to avoid difficult, embarrassing, inconvenient, time-consuming situations as well as those activities which make no sense to them. (Thus, they fail to make some assigned observations or to ask some of the interview questions.) At the same time they want to give the right impression to their superiors—at least right enough so that their material will be accepted and they will be kept on the job. (Thus, they modify or fabricate portions of the reports in order to give the boss what he *seems* to want.) They do not want to "look stupid" by asking too many questions, so they are likely to make a stab at what they think the boss wants—e.g., make a guess at a coding category rather than having it resolved through channels.

Even those who start out with the notion that this is an important piece of work which they must do right will succumb to the hired-hand mentality when they realize that their suggestions and criticisms are ignored, that their assignment does not allow for any imagination or creativity, that they will receive no credit for the final product, in short, that they have been hired to do somebody else's dirty work. When this realization has sunk in, they will no longer bother to be careful or accurate or precise. They will cut corners to save time and energy. They will fake parts of their reporting. They will not put themselves out for something in which they have no stake except in so far as extrinsic pressures force them to. Case No. I is an excerpt from the statement of a research worker who started out with enthusiasm and hard work and ended with sloppy work and cheating when she could no longer escape the fact that she was a mere flunky expected to do her duty whether or not it was meaningful. The coders in Case II soon gave up any effort to resolve the ambiguities of their coding operation and followed the easiest path acceptable to their supervisor. In this case, the supervisor himself made little effort to direct the data-processing toward supplying answers to meaningful research issues. We must remember that in many research operations the supervisors and directors themselves are hired hands carrying out the requests of a client or superior as expeditiously as possible.

Many of the actions of hired hand researchers are strikingly analogous to restrictive practices of factory operatives. Interviewers who limit probing and observers who limit interaction recording are behaving like workers applying "quota restriction," and with interacting hired hands informal agreements may be reached on the extent of such restrictions. To fabricate portions of a report is a form of goldbricking. The collusion on the reliability check reported in Case I is strikingly similar to the workers' plot to mislead the time-study department. Such similarities are no accident. The relationship of the hired hand to the product and the process of production is the same in each case. The product is not "his." The production process gives him little or no opportunity to express any intrinsic interest he may have in the product. He will sooner or later fall into a pattern of carrying out his work with a minimum of effort, inconvenience, and embarrassment—doing just enough so that his product will get by. If he is part of a large and complex operation where his immediate superiors are also hired

---

[1] Donald Roy, "Quota Restriction and Goldbricking in a Machine Shop," *American Journal of Sociology*, 57 (March 1952), pp. 427–442.

POOR ORIGINAL COPY - BEST
AVAILABLE AT TIME FILMED

hands with no intrinsic interest in the product and where the final authority may be distant and even amorphous, quality control of the product will be mechanical and the minimal effort that will get by can soon be learned and easily applied. The factory production situation has at least one ultimate limitation on the more extreme deviations of the hired hands: the final product must "work" reasonably well in a substantial proportion of cases. In social science research, on the other hand, the product is usually so ambiguous and the field of study so lacking in standards of performance, that it is difficult for anyone to say whether it "works" or not.

What is more important is the effect of the hired hand mentality on the *nature* of the product. Workmen not only turn out less than they could if it were in their interest to maximize production, but often produce shoddy and even dangerous products.[2] In the case of research, the inefficiency of hired hands not only causes a study to take longer or cost more money, but is likely to introduce much dubious data and interpretations into the process of analysis. Our mass production industrial system has opted to sacrifice individual efficiency and product quality for the advantages of a rationalized division of labor. The same approach has been applied to much of our larger scale scientific research and the results, in my opinion, have been much more disastrous than they are in industrial production with little of the compensating advantages.

When the tasks of a research project are split up into small pieces to be assigned to hired hands, none of these data-collectors and processors will ever understand all the complexities and subtleties of the research issues in the same way as the person who conceived of the study. No amount of "training" can take the place of the gradual development of research interests and formulations on the part of the planner. Since the director often cannot be sure what conceptions of the issues the hired hands have as a result of his explanations and "training," he must make dubious guesses about the meaning of much of the data they return to him. If he attempts to deal with this difficulty by narrowly defining the permissible behavior of each hired hand (e.g., demand that all questions on a schedule be asked in a set wording), he merely increases the alienation of the hired hand from his work and thus increases the likelihood of cutting corners and cheating. As he gains in quantity of data, he loses in validity and meaningfulness.[3]

I do not want to give the impression that the hired hand mentality with its attendant difficulties is simply a characteristic of the large-scale on-going research organization. We may find it at all size levels, including the academic man hiring a single student to do his research chores. The argument may be advanced that assignment of specified tasks by the director of a study is essential to getting the job done in the manner that he wants it done. My answer is that such assignments are often not effectively carried out and it is misleading to assume that they are.

Let me illustrate this point. A researcher wants to do a study of the operation of a given institution. He has some definite notion of what aspects of behavior of the institutional personnel he wants information about and he has some ideas about the manner in which he will go about analysing and interpreting these behaviors. He finds it possible and useful to engage four trained and interested assistants. Let me outline two ways the study might be conducted:

A. Through a series of discussions, general agreement is reached about the nature of the study and the manner in which it might be conducted. Some division of labor is agreed upon in these discussions. However, none of the field workers is held to any particular tasks or foci of interest. Each is allowed to pursue his data-collection as he thinks best within the larger framework, although the field workers exchange information frequently and make new agreements so that they can benefit from each other's experience.

B. The director divides up the data-collection and processing in a logical manner and assigns a portion to each of the assistants. Each field worker is instructed to obtain information in all the areas assigned to him and to work in a prescribed manner so that his information will be directly comparable to that of the others. The director may use a procedural check such as having each assistant write a report covering given issues or areas at regular intervals.

Which is the preferred approach? Judging from my reading of social science journals, most research directors would say Method B is to be preferred. Method A, they would maintain, produces information on subjects, issues, or events from one field worker which is not directly comparable to that collected by another field worker. They would also object that if each field worker is permitted to follow his own inclinations even in part, the total study will suffer from large gaps. These accusations are quite true—and, I would add, are an inevitable result of dividing a research project among a number of people. What I disagree with, however, is the assumption that Method B would not suffer from these defects (if indeed, they should be regarded as defects.) It is assumed that the assistants in Method B are actually carrying out their assigned tasks in the manner specified. In line with my earlier discussion of the behavior of hired hands, I would consider this highly unlikely. If the information produced by these assistants is indeed closely comparable, it would most likely be because they had reached an agreement on how to restrict production. And, whether the study is carried out by Method A or by Method B, gaps will occur. The difference is that the director of Study A—assuming he had succeeded in making his assistants into collaborating colleagues—would at least know where the gaps are. The director of Study B would have gaps without knowing where they are—or indeed, that they exist—because they have been covered over by the fabrications of his alienated assistants.

It is ironic that established researchers do not ascribe the same motivating forces to their subordinates as they do to themselves. For many years research scientists have been confronting those who pay their salaries and give them their grants with the argument that a scientist can do good research only when he has the freedom to follow his ideas in whatever way seems best. They have been so successful with this argument that university administrations and research organization directorates rarely attempt to dictate—or even suggest—problems or procedures to a researcher on their staff, and the more prominent granting agencies write contracts with almost no strings attached as to the way in which the study will be conducted. Yet research directors fail to apply this same principle to those they hire to carry out data-collection and processing. The hired assistant's desire to participate in the task and the creative contribution he might make is ignored with the result that the assistants' creativity is applied instead to covertly changing the nature of the task.

---

[2] I want to emphasize once again that in a business setting, supervisors and executives, as well as production line workmen, participate in aspects of the hired hand mentality. None of them may have an intrinsic interest in the quality of the product. (See, for example, Melville Dalton, *Men Who Manage*, New York: John Wiley and Sons, Inc., 1959, especially Chapters 7, 8, and 9.) The same is the case in much large-scale research.

[3] In this discussion I am assuming there is some one (or a small group of colleagues) who has initially formulated the research problem or area of concern because of intrinsic interest and curiosity. In much of our social science research, we do not have even this saving grace and the research is formulated and carried out for various "political" reasons. In such cases, we cannot count on having anyone interested enough to try to turn the accumulations of data into a meaningful explanatory statement.

There has been very little discussion in our journals and our books on research methods on the relationship of the hired hand to the data collected. Whatever discussion there *has* been can be found in the survey interview field where there have been some studies of the effect of such demographic factors as age, sex, and race, sometimes measured personality traits, on "interviewer bias." The nature of the interviewer's status in a research organization is seldom discussed in print. The problem of interviewer cheating, although a common subject of informal gossip, is seldom dealt with openly as a serious problem. When Leo Crespi published an article twenty years ago in which he expressed the worry that cheating was seriously affecting the validity of much survey data,[4] those who responded (mostly survey organization executives) stated reassuringly that few interviewers cheated and that they had pretty effective ways of controlling those who did.[5] If the analysis offered in this paper is correct, the first part of this reassurance is almost certainly wrong. The low-level flunky position which most interviewers occupy in survey organizations[6] should lead us to expect widespread deviations from assigned tasks. The survey executives who responded give no convincing evidence to the contrary. As for the second part of the assertion, their descriptions of their control measures indicate that they can hope to block only the cruder, more obvious, and repeated forms of cheating. The postal card follow-up will catch the interviewer who does not bother to contact his respondents at all. Spot-check follow-up interviewing may eventually catch the interviewer who makes contacts, but fabricates demographic data (to fill a quota sample) or completes only part of the interview and fills in the rest in a stereotyped manner later on. (Even here, many of his interviews may be used before he is detected.) However, from the cases of hired hand interviewing which I am familiar with, I would say such crude cheating is not the most common form of cutting corners on the job. Far more common is the kind found in Case III where the interviewer makes his contact, obtains a fairly complete interview, but leaves partial gaps here and there because he found it time-consuming, embarrassing, or troublesome, felt threatened by the respondent, or simply felt uncertain about how the study director wanted certain lines of questioning developed. With a little imagination, such gaps can be filled in later on in a way that is very unlikely to be detected in a follow-up interview. If, for example, a supervisor in Case III had returned to the respondents and asked them whether the "five reasons" listed on their interview form were accurate reflections of their opinion, probably most would have said yes, and the few who objected to one or two of the reasons could have been dismissed as the degree of change that one expects on re-interview.[7]

Some gimmicks for catching cheaters may even put the finger on the wrong person. Thus, one approach to detecting cheating is to compare the data of each interviewer to the group averages and to assume that if one deviates markedly from the group, he is cheating or doing his work improperly. This reasoning assumes that cheating is exceptional and will stand out from the crowd. I have already suggested that the opposite is often the case. Therefore, if the cheaters are working in the same direction (which is readily possible if they have reached an informal agreement or if the question is of such a nature as to suggest distortion in a given direction), it is the "honest" person who will deviate. In the study alluded to in Case III, for example, one of the interviewers always left spaces open on the "five reasons" item. At one point the director reprimanded him for not obtaining five responses "like the rest of the interviewers." The director preferred to believe that this man was not doing his job right than to believe that all the rest were making up responses.

Large survey organizations have at least made some attempts to control the cruder forms of cheating. In most studies using hired hands, even this limited control is absent. The academic man with one or a few assistants, the research organization study director with one or a few small projects, usually has no routine way of checking on the work of his assistants. If he duplicates much of their work or supervises them very closely, he may as well dispense with their services. If he gives them assignments without checking on them closely, he is in effect assuming that they are conducting their assignment more or less as directed and is accepting their products at face value. This assumption, I assert, is a dubious one. And since it is a common practice nowadays to farm out much of one's research work—quite often to accumulate research grants only to hire others to do the bulk of the work—the dubious nature of hired hand research is a widespread problem in small as well as large scale research, in surveys, in direct observation, and in various forms of data processing.

I do not want to suggest, however, that the major failure of hired hand research is the lack of control of cheating. Rather, the very fact that we are placed in a position of having to think up gimmicks to detect cheating is in itself an admission of failure. It means that we are relying for an important part of our research operation on people who have no concern for the outcome of the study. Such persons cannot have the kind of understanding of the data-collection or data-processing procedures which can come only with working out problems in which the researcher has an intrinsic interest and has gone through a process of formulating research questions and relevant ways of collecting and processing data.

I can hear the objection that much social science cannot be done without hired hands. But we should at least be aware of the doubtful nature of some of the information collected in this way and construct our data-collection and processing in such a way as to reduce the encouragement of cheating and restriction of production as much as possible. (See Crespi's list of "ballot demoralizers." [8]) More important, however, I believe the need for hired hands has been greatly exaggerated. Why, for example, must we so often have large samples? The large sample is frequently a contrivance for controlling various kinds of "errors" (including the "error" introduced by unreliable hired hands). But if the study were done on a much smaller sample by one person or several colleagues who formulated their own study and conducted it entirely by themselves, much of this error would not enter in the first place. Isn't a sample of fifty which yields data in which we can have a high degree of confidence more useful than a sample of five thousand where we must remain doubtful about what it is that we have collected? Often a large-scale study tries to do too much at one time and so ends up as a hodge-podge affair with no integration of ideas or information ever taking place because it is, in effect, *nobody's* study. How often have you read the report of a massive study expending large amounts of money and employing large numbers of people where you were disappointed at the paucity of the results, especially when compared to a far smaller project on a similar issue conducted entirely by one or a few people?

---

[4] Leo Crespi, "The Cheater Problem in Polling," *Public Opinion Quarterly*, Winter 1945–1946, pp. 431–445.

[5] "Survey on Problems of Interviewer Cheating," *International Journal of Opinion and Attitude Research*, 1 (1947), pp. 93–107.

[6] Julius A. Roth, "The Status of Interviewing," *The Midwest Sociologist*, 19 (December 1956), pp. 8–11.

[7] I have even heard the argument that it makes no difference if perceptive interviewers make up parts of the interview responses with the help of information from other responses because their fabrications will usually closely approximate what the subject would have said if he could have been prompted to answer. But if we accept this argument, a large portion of the interview should have been eliminated to begin with. It means we already claim to know the nature of some of the relationships which the study is purportedly investigating.

Let me repeat that I am not singling out large-scale operations as the only villains. The current structure of professional careers is such that often small studies are turned over to hired hands. We tend to be rated on how many studies we can carry on at the same time rather than on how thoroughly and carefully we can carry through a given line of research. Soon we find that we do not have time for all of the projects we have become involved in and must turn some over to others of lower professional status. This might not be so bad if we were willing to turn over the research work wholeheartedly. We might simply act as entrepreneurs to funnel funds to others and to provide them with appropriate clearance and an entré to research settings. We can then leave the specific formulation of the problem and procedure (and the credit for doing the work) to the person we have helped out. Such is often done, of course. However, there are many instances in which the senior researcher believes those he has hired cannot be trusted to formulate their own plans, or professional career competition convinces him that he cannot "afford" to give up any of his studies to others. In such cases he is likely to maintain a semblance of control by mechanically structuring a research plan and making assigaments to his assistants. This, as I have indicated, is the way to the hired hand mentality with its attendant distortions of research data.

What is a hired hand? So far I have been talking as if I knew and as if the hired hand could readily be distinguished from one who is not. This, of course, is not true. The issue is a complex one and information on it is, by its very nature, not very accessible. It is a crucial question which deserves study in its own right as part of the more general study of the process of "doing research."

Let me attempt a crude characterization of hired hand research, a characterization which hopefully will be greatly refined and perhaps reformulated with further study. A hired hand is a person who feels that he has no stake in the research that he is working on, that he is simply expected to carry out assigned tasks and turn in results which will "pass inspection." Of course, a hired assistant may not start out with the hired hand mentality, but may develop it if he finds that his talents for creativity are not called upon and that his suggestions and efforts at active participation are ignored.

From specific examples from the research world and by analogy from research on hired hands in other occupational spheres, I am convinced that research tasks carried out by hired hands are characterized, not rarely or occasionally, but *typically*, by restricted production, failure to carry out portions of the task, avoidance of the more unpleasant or difficult aspects of the research, and outright cheating. The results of research done in part or wholly by hired hands should be viewed as a dubious source for information about specific aspects of our social life or for the raw material for developing broader generalizations.

Of course, this leaves open the question of what constitutes a "stake in the research" and how one avoids or reduces the hired hand mentality. Again, I have no specific answers and hope that issue will receive much more attention than it has up to now. A stake may mean different things in various circumstances. For graduate students, a chance to share in planning and in writing and publication may often be important. For interviewers or field workers, the determination of the details of their procedure may be crucial. In an applied setting, the responsibility for the practical consequences of the research findings may be most important.[9]

It would also be worthwhile to examine the conditions which make for hired hand research. Here again, I have little specific to say and this subject, too, needs much more investigation. However, I will suggest a few factors I consider important.

Size: Hired hands can be found in research staffs of all sizes from one on up. However, it is clear that when a very small number of researchers are working together, there is a greater possibility of developing a true colleagueship in which each will be able to formulate some of his own ideas and put them into action. The larger the group, the more difficult this becomes until the point is probably reached where it is virtually impossible, and the organization must be run on the basis of hierarchical staff relations with the lower echelons almost inevitably becoming hired hands.

Subordination: If some members of the research group are distinctly subordinate to others in a given organizational hierarchy or in general social status, it will be more difficult to develop a true colleague working relationship than if their status were more closely equal. The subordinate may hesitate to advance his ideas; the superordinate might be loath to admit that his lower-level co-worker be entitled to inject his ideas into the plans. Formal super-subordinate relationships can of course be muted and sometimes completely overcome in the course of personal contact, but certainly this is an initial, and sometimes permanent, basis for establishing hired hand status.

Adherence to Rigid Plans: If a researcher believes that good research can be done only if a detailed plan of data-collection, processing, and analysis is established in advance and adhered to throughout, he has laid the basis for hired hand research if he makes use of assistance from others who have not participated in the original plan. Sticking to a pre-formed plan means that others cannot openly introduce variations which may make the study more meaningful for them. Any creativity they apply will be of a surreptitious nature.

---

[8] Leo Crespi, *op. cit.*, pp. 437-439.

[9] The "human relations in industry" movement has given us some useful suggestions about the circumstances which alienate workers and executives, and also ways in which industrial employees may be given a real stake in their jobs. See, for example, Douglas McGregor, *The Human Side of Enterprise*, New York: McGraw-Hill Book Co., 1960, Part 2.

In their research methods texts, our students are told a great deal about the mechanics of research technique and little about the social process of researching. What little is said on the latter score consists largely of Pollyannaish statements about morale, honesty, and "proper motivation." It should be noted that appeals to morality and patriotism never reduced goldbricking and restriction of production in industry, even during the time of a world war. There is no reason to believe that analogous appeals to interviewers, graduate students, research assistants, and others who serve as hired hands will be any more effective. If we want to avoid the hired hand mentality, we must stop using people as hired hands.

Glaser and Strauss state that we regularly "discount" aspects of many, if not most, of all scientific analyses we read because we consider the research design onesided, believe that it does not fit the social structure to which it was generalized, or that it does not fit in with our observations in an area where we have had considerable experience.[10]

I would like to suggest another area in which we might consistently apply the "discounting process." When reading a research report, we should pay close attention to the description of how the data was collected, processed, analyzed, interpreted, and written up with an eye to determining what part, if any, was played by hired hands. This will often be a difficult and highly tentative judgement, requiring much reading between the lines with the help of our knowledge of how our colleagues and we ourselves often operate. However, we can get hints from such things as the size of the staff, the nature of the relationship of the staff members, the manner in which the research plans were developed and applied, the organizational setting in which the research was done, mention made of assignment of tasks, and so on. If there is good reason to believe that significant parts of the research has been carried out by hired hands, this would, in my opinion, be a reason for discounting much or all of the results of the study.

---

[10] Barney Glaser and Anselm L. Strauss, "Discovery of Substantive Theory: A Basic Strategy Underlying Qualitative Research," *American Behavioral Scientist*, 8 (February 1965), pp. 5-12.

CHAPTER TWO

CLASS NOTES

I

In teaching these materials to law students, we have often encountered a difficult problem. Students have frequently not had the training which will enable them to formulate an hypothesis clearly and well. It is not an esoteric skill, of course, and so it may be surprising that such a task initially poses so many difficulties. But the problem does exist, and must be overcome if the student is to learn how to fit social science data to the legal process.

A large share of the problem, we are convinced, is due to the frequent lack of familiarity with the meaning of the term relationship, especially when it involves quantitative material which is best expressed in graphic or tabular form. Later in this book we will go into the meaning of relationship in a good deal of detail. However, some understanding of the term is necessary even in the very first stages of considering the nature of inquiry in the social sciences; and the following notes should be viewed as no more than a brief introduction to a concept.

II

We have heard it said that children using one brand of toothpaste get 24% fewer cavities than children using another brand. In baseball, some pitchers do better in night games than others who seem to do better during the day. In some states population is increasing; in others it is decreasing. We have been told that people who smoke are more likely to die of lung cancer (and other diseases) than people who don't. We often hear that young drivers have more automobile accidents than older drivers. Some people say that women who drive cars have more accidents than men.

Among these and thousands of other seemingly related facts, it has been found helpful to be able to state the relationships in numerical, or quantitative, terms--more and less, higher and lower, rising and falling. Businessmen frequently want to know which products will be used often and which less often, and the answer may depend on what kind of people use them; candidates for public office sometimes want to know what sorts of people are most likely to vote for them; school board members--and voters and house buyers--need to know in which areas of a city population is likely to increase.

Stating the relationships between two things in quantitative terms is not always simple, although it sometimes appears to be. In the following pages, we are going to consider some of the ways in which relationships can be stated quantitatively, some of the problems that arise in formulating such statements, and some of the traps to avoid.

TABLE I

YEARS OF FORMAL EDUCATION AND ANNUAL INCOME OF A SAMPLE OF 50 INDIVIDUALS

Identifying Number for Individuals	Years of Formal Education	Annual Income
1	9	$3900
2	12	$7300
3	11	$5600
4	8	$3200
5	12	$6700
6	12	$5800
7	11	$6900
8	11	$6200
9	10	$4700
10	9	$4200
11	11	$6400
12	11	$5300
13	10	$6600
14	10	$5800
15	11	$5100
16	11	$5300
17	8	$4100
18	8	$4300
19	11	$4900
20	10	$6300
21	11	$6300
22	10	$6300
23	10	$5600
24	8	$2900
25	9	$5100
26	11	$5400
27	11	$4900
28	12	$6700
29	10	$5600
30	10	$6800
31	11	$5400
32	10	$6400
33	8	$4000
34	9	$4500
35	10	$6200
36	10	$4600
37	11	$5900
38	10	$5700
39	10	$5600
40	8	$3800

TABLE I (Continued)

Identifying Number for Individuals	Years of Formal Education	Annual Income
41	10	$4300
42	8	$3900
43	10	$4900
44	12	$5900
45	10	$4900
46	9	$5000
47	12	$6200
48	8	$4600
49	10	$5900
50	12	$5900

Let's start with a fairly easy example of how a relationship can be expressed. What is the relationship between the amount of formal education that people have and the income these people earn? You can undoubtedly guess in a general way what the relationship is.

In Table I you will find a list of a sample of fifty people, with each person assigned an identifying number. For each individual information is provided concerning his years of formal education and annual income. With this information, you can express the relationship in a kind of picture called a scatter diagram. (You may have come across scatter diagrams in your mathematical courses.) The diagram can be made on a sheet of graph paper. Place a dot on the paper at the point where the line for annual income (indicated along the horizontal axis) crosses the line for years of formal education (indicated along the vertical axis). Thus, the dot representing income and education for individual #1 will be located where the vertical line that represents $3,900 meets the horizontal line that represents 9 years. If the figure you are plotting is not shown on the graph, you can estimate its position between figures that are shown.

The entire set of dots, scattered across the graph paper, is a picture or diagram of the relationship between education and income for this sample of 50 individuals. By looking at the pattern the dots make on the paper, you can tell whether your guess about the relationship was right.

TABLE 2

NATIVITY AND EDUCATION IN THE UNITED STATES
(by geographical division)

Geographical Division	% of Population Foreign-Born	% of Population Uneducated
1. New England	10.1	55.5
2. Middle Atlantic	10.3	60.0
3. East North Central	5.2	58.8
4. West North Central	2.6	57.1
5. South Atlantic	2.1	63.4
6. East South Central	0.5	70.5
7. West South Central	2.1	62.8
8. Mountain	3.7	50.3
9. Pacific	7.9	48.9

In general, individuals with low education have low income, and individuals with high education have high income.

When two such phenomena as education and income vary together in this way, we say they are correlated, or that there is a correlation between them.

Let's turn to another example involving the relationship between the percentage of people in a state who were born in a foreign country and the percentage of people in the state who have not had a high school education or its equivalent. (For convenience we will use the term educated in this study to mean high school education or more; uneducated, less than high school education.)

The states have been divided into nine geographical blocks with the percentage of foreign-born and the percentage of uneducated given for each division. The data are given in Table 2. Construct the scatter diagram as you did the one before, and by looking at the pattern of dots decide whether your guess about the relationship of the variables was right. It may help you to decide if you try to draw a single straight line that comes as close as possible to touching all the dots. A line of this kind is called a regression line. (There are mathematical ways of figuring out exactly where to draw a regression line in any particular scatter diagram.)

The kind of correlation that your diagram shows between the percentage of foreign-born and the percentage of uneducated may surprise you. If you plotted the diagram accurately, it shows that a low percentage of foreign-born often goes with a high percentage of uneducated. There is an important lesson here: even things that are "obviously true," as we too often say, sometimes turn out not to be true when the evidence is examined. One advantage of stating relationships in precise quantitative terms is that it enables us to examine evidence carefully.

The kind of correlation in which high numbers along one axis tend to go together with low numbers on the other axis is called a negative correlation. When a correlation is negative, the regression line will slant from the upper left towards the lower right. In a positive correlation, such as that between education and income, the higher numbers along one axis tend to go together with higher numbers on the other axis. Then, of course, the regression line will slant from the lower left toward the upper right. (You may wish to draw a regression line on your first scatter diagram.)

TABLE 3

DIMENSIONS OF 10 RECTANGLES

RECTANGLES	HEIGHT	WIDTH
1	2.5	1
2	5	6
3	4	4
4	3.75	3.5
5	4.75	5.5
6	4.5	5
7	3	2
8	4.25	4.50
9	6.5	9
10	6	8

III

When you tried to draw regression lines on the previous scatter diagrams, you found that no single straight line would go through all the dots. If the dots in a scatter diagram are so arranged that a single straight line will pass through all of them, the correlation is said to be perfect. The data in Table 3, the dimensions of a set of rectangles, give you a chance to see what a perfect correlation looks like. (The rectangles had to be made up especially because no data that would show a perfect correlation are available from sociological sources. In the social sciences, perfect correlations are never encountered.)

If we plot the height and the width of each rectangle, we have a picture of a perfect correlation. If you place the dots accurately, you will find that you can draw a straight line that goes through all the dots. Is the correlation between length and width positive or negative.?

One of the advantages of a scatter diagram and its regression line is that it enables us to predict one set of figures from another set. If we know the height of one of the rectangles, in this particular set of rectangles, we can predict its width.

Remember, though, however possible mathematically, perfect correlations are never found in the social sciences. Social science is concerned

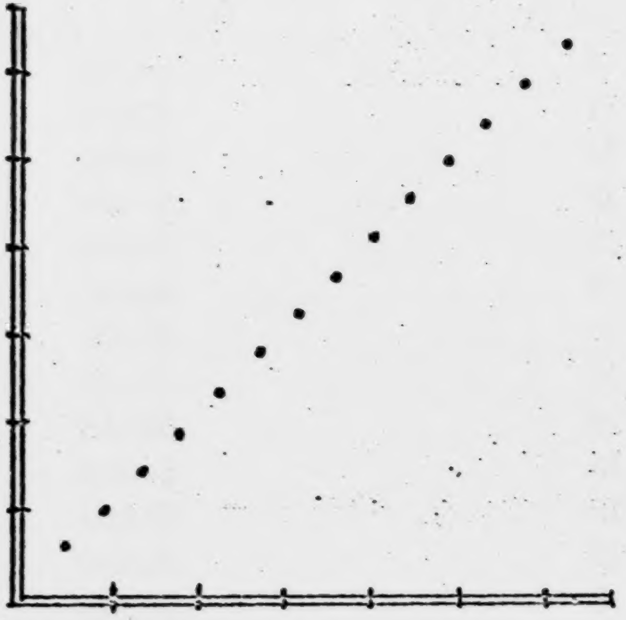


Figure 2. Perfect positive correlation

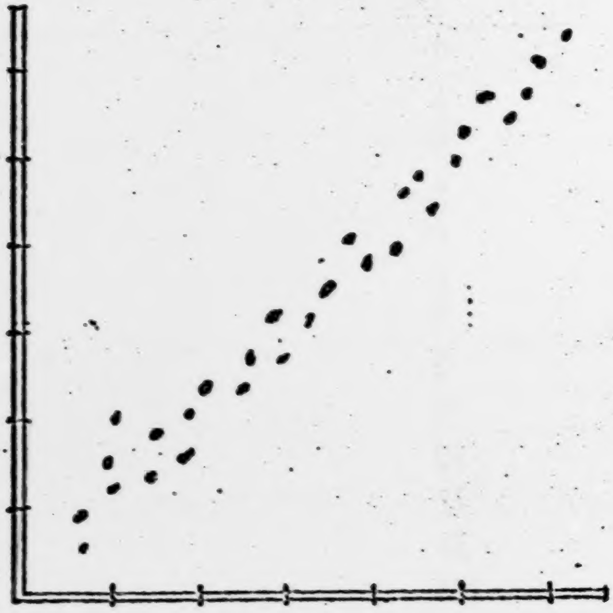


Figure 3. High positive correlation

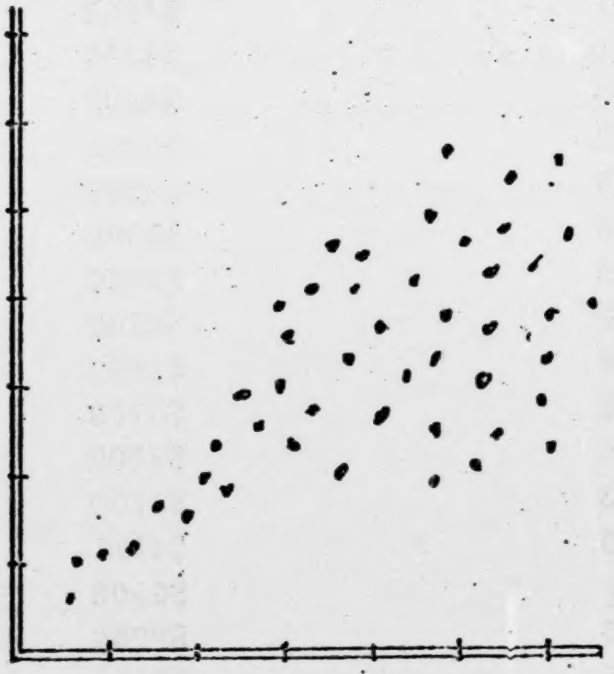


Figure 4. Low positive correlation

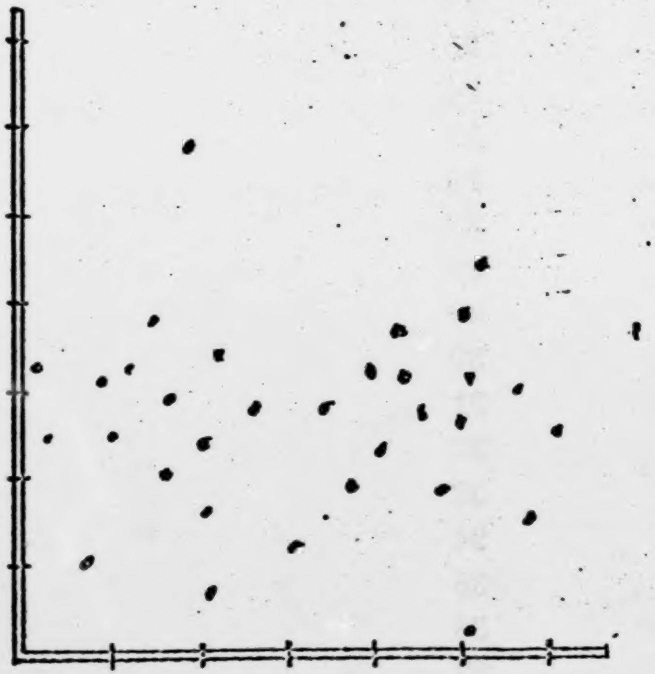


Figure 5. No correlation

TABLE 4

EDUCATION AND INCOME OF 40 FATHERS

Case	Education (Years)	Income ($ per year)
1	8	$3200
2	17	$6600
3	18	$8500
4	12	$3100
5	12	$3600
6	19	$9500
7	10	$4200
8	16	$5500
9	10	$5000
10	15	$4500
11	19	$7000
12	12	$5400
13	8	$3300
14	8	$3700
15	7	$2400
16	6	$2800
17	5	$1700
18	16	$7000
19	12	$6000
20	12	$4300
21	12	$4400
22	11	$4400
23	11	$4300
24	10	$4200
25	10	$3300
26	13	$3900
27	9	$4500
28	9	$5000
29	13	$4700
30	14	$4900
31	14	$5700
32	14	$5800
33	15	$6200
34	9	$4100
35	16	$6500
36	16	$8700
37	16	$7500
38	11	$3000
39	6	$ 900
40	7	$1200

with human beings, and the characteristics of human beings just do not seem to be perfectly correlated. Two sets of social facts, if they are related at all, have up to this time been found to be somewhat related. The first two scatter diagrams you made are typical demonstrations of the less than perfect correlations that obtain among human characteristics. Figures 2, 3, 4, and 5 also illustrate possible degrees; that is, of stating just how far from (or near to) a perfect correlation any given relationship comes. (We will describe one of the ways to measure this a little farther on.)

The scatter diagram derived from Table 1 contained a large amount of information. You probably found that making the diagram was time-consuming and tedious work. Often it is helpful to be able to boil a lot of information down to a few relatively simple figures. One reason is that highly detailed figures, especially in the social sciences, cannot always be trusted. If one man tells us that his income last year was $10,327.18 and another man says his income was $10,317.60, we wouldn't--and shouldn't--be certain that there was a difference worth noticing between their incomes; we'd feel more confident in our observation of their incomes if we said: "They both made more than $10,000." Another reason for wanting to condense information is that it is hard to look at many items of information and even harder, unless you have a computer handy, to remember them and deal with them at the same time.

In order to see how an unwieldy number of observations can be condensed, we will once more use a scatter diagram. Table 4 provides a list of numbered students in a high school class, showing the years of education and the incomes of their fathers. Construct a scatter diagram of the fathers' education and incomes, using the piece of graph paper marked for that purpose. As you would expect, the correlation is positive--the higher the fathers' education, the higher their income. Although the figures you are using have been made up for illustrative purposes, the relationship of any individual's education and income is pretty generally positive. Draw an approximate regression line on your plotted graph for the 40 fathers' incomes and years of education.

One way of condensing the information in this scatter diagram is by grouping the students' fathers into income brackets. Just as the states were grouped into nine dividions in order to simplify the task of plotting the data on nativity and education, so we can group the fathers into those earning less than $1,000 a year, those earning from $1,000 to $1,999, and so on. Then we can measure those fathers in each income bracket according to the mean[1] years of education for the group taken collectively. We then put on the graph, just as we did for the 40 incomes and years of education, a dot where the mean income for any of our new income groups crosses the mean years of education of the group. Thus we can reduce the 40 dots of our first diagram to only 10 dots. ($9,000 - $9,999 is the highest bracket for our figures.)

---

[1] Total years of education for all fathers in any income bracket divided by the number of fathers in the bracket.

Carry out the needed calculations and plot the results. It will probably help if you start by drawing vertical lines on your scatter diagram at the points on the horizontal axis where the divisions between the income brackets fall (that is, at each $1,000 of income). Does your second grouping of figures give you the same general correlation that your first scatter diagram gave?

You can, of course, condense your original information (incomes and education of 40 fathers) by reversing the order of condensation that you just followed. You should do so, because it will show you something interesting and important. Begin, then, by clustering the fathers according to education brackets; then calculate the mean income and the mean education for each bracket. Now put your dots (better, put x's to distinguish the two plottings) to represent the mean education and mean income for the fathers in each individual bracket. You will find that the x's are not where the dots are. It follows that the regression line will not be the same, either. Such a difference is a signal that you have a less-than-perfect correlation. It tells you that to predict income from a knowledge of education you need a formula somewhat different from the one you used to predict education from a knowledge of income.

There is no reason why we have to stick to intervals of $1,000 and one year. We could just as well use $2,000 and 2 years or $3,000 and 5 years or some other combination. If we wanted to boil down all the information into the simplest form that would still show whether there was a relationship between income and education, we could use just two intervals along each axis.

What should these two intervals be? That question is not always an easy one to answer. You will usually arrive at an appropriate answer if you make each of your two intervals contain about the same number of persons. In the present example, we will use as the dividing lines for the intervals the median[2] income of employed men in the United States, which is about $4,700; and the median education of all adult males in the United States, which is about 10 years. With these intervals, we can count on getting about half the fathers into each of the two income categories and about half into each of the two education categories--if, that is, the fathers are representative of the United States population as a whole.

---

[2] The median is another average (or "measure of central tendency") which is somewhat different from the mean, as will be explained in the readings. The median is defined as that number which is smaller than half the numbers in a set of numbers and larger than half the numbers.

In this new condensation our two intervals for income will be (1) $0 - 4,699 and (2) $4,700 and everything more than that; our intervals for education will be (1) anything less than 10 years, and (2) 10 years and more. On the scatter diagram you have been working with, draw heavy lines to mark these intervals; they will divide your diagram into four squares. It is not necessary to calculate the means for each of these squares. There is another and easier way of summarizing the information in the diagram; simply count the number of fathers shown in each square. You will then know how many of them are in the upper bracket of both income and education, how many are in the lower bracket of both, and how many are in the upper bracket of one and the lower bracket of the other. Make these counts, and enter the numbers in the table following. The four squares in this table correspond to the four squares in your diagram; and education of 10 years or more is called high, and an income of $4,700 a year or more is called high:[3]

---

[3] Because the terms high and low do not have any fixed quantitative meaning, we can use them to stand for anything we choose. However, when you see the words in a table, be sure you know exactly what meaning they have been given.

TABLE 5

Father's Education	Father's Income	
	Low	High
High		
Low		

Since a table of this kind has four figures in it, it is often called a four-fold table. (Each of the squares is a cell.) The sum of the figures that you put in the cells should be 40; that is how many fathers we have information on.

A large amount of the detail that you could observe in your scatter diagram has been lost in this table. But the loss of detail is at least partly offset by a gain in simplicity; you need to look at only four numbers to get a general picture of the relationship.

To get a more exact picture of the relationship, notice that a positive correlation means that a relatively large proportion of the fathers will show up either in both the high cells (high education and high income) or in both the low cells. You will remember that a positive correlation was defined as one in which a high position on one variable goes with a high position on the other. On the other hand, when the correlation is negative, the units (fathers, in this table) will bunch up in the cells that are high for one measurement and low for the other. This unvarying pattern of distribution led to a simple way to calculate the degree of correlation between two sets of measurements. Multiply together the number of units in the high&high[4] cell and the number of units in the low&low cell. From that product subtract the product of the number of units in the high&low cell and the number of units in the low&high cell. Express the difference between the products as a proportion of (a ratio to) the sum of the same products. Here is the calculation stated in a formula:

$$\frac{(\text{high\&high} \times \text{low\&low}) - (\text{high\&low} \times \text{low\&high})}{(\text{high\&high} \times \text{low\&low}) + (\text{high\&low} \times \text{low\&high})}$$

If all the units were in the high&high and low&low cells, the formula would give you a positive 1; you would have a perfect negative correlation. Other degrees of correlation will be fractions (often expressed in decimals) somewhere in between. The higher the positive correlation between two statements, the nearer the ratio will be to positive 1; the higher the negative correlation, the nearer the ratio will be to negative 1. Calculate the ratio for the relationship between fathers' education and fathers' income (Table 5). Would you say it is a high correlation or a low one?[5]

---

[4] Note that this term and the three similar terms following are written as one word.

It is not necessary to start with a scatter diagram every time we want to analyse data by a four-fold table. We can simply classify each unit (or rather, each observation we make about a unit) according to each of two variables, put the unit in the appropriate cell, and then add up all the observations in each cell. For example, here is a table in which all the adults in the United States in 1960 have been classified according to (a) whether they completed high school or left school before completion and (b) whether they were born in the United States or in a foreign country:

---

[5] There are many different ways of measuring the degree of correlation between two variables. It takes an expert in statistics to know which is the proper measure to use in a particular case. The measure we have just described is called Q, in honor of the Belgian statistician Quetelet. Another measure often used is the *phi coefficient*. The formula for the phi coefficient is

$$\frac{BC - AD}{\sqrt{(A + B)\ (C + D)\ (A + C)\ (B + D)}}$$

where B and C are the high&high and low&low cells and A and D are the two others.

TABLE 6

	NATIVITY	
Education	Foreign-born	U.S.-born
	(by thousands of persons aged 25 or more)	
Educated	2,345	48,043
Uneducated	6,935	68,508

We worked with information about this relationship earlier, but we used then information about geographical divisions (more precisely, about proportions of individuals according to the part of the country in which they live); here we are using information about each individual, without regard to the region of the country in which he lives.

One point about Table 6 should be explained before we go on. Nativity is a different sort of variable from income and education. Income and education come in many different amounts, or qualities; they are therefore called quantitative variables. Nativity, by contrast, does not come in different amounts; you are either born in a certain country or you are not. Nativity is a qualitative variable; it comes in different kinds, or qualities, rather than in different amounts. You can be more educated or less educated, but while people are born in many different countries, whatever country they were born in, they were either born there or they were not.

An important qualitative variable in sociological investigations is sex; all persons are either male or female, and everyone can be placed in one of the two categories. When dealing with the variable of sex and all other qualitative variables, four-fold tables become particularly common. For example, in a study of factors that affect whether people vote in an election, you would be forced to use a four-fold table if you wanted to know whether sex was one of the factors. Every person you studied would have to be classified as either male or female and as either voting or not voting.

It is true that any quantitative variable can be turned into what looks like a qualitative variable. We converted income and education from quantitative to qualitative measurements in Table 5. A person, we said in the table, either earned less than $4,700 or he did not; he had completed high school or he had not. However, there we could still speak--and did speak-- of high and low income and education. Since truly qualitative variables do not come in amounts, we cannot use the terms high and low in connection with them. But we can still speak of positive and negative correlations, as long as we are careful to make clear what we mean. For example, if being born in the United

States tended to go with a high school education, we could say there was a positive correlation between being native-born and being educated, or a negative correlation between being native-born and uneducated.

Now let's find out whether the correlation between being educated and being born in the United States is positive or not. Calculate Q for Table 6, taking, for the purpose of calculation, the cells labeled educated and the cells labeled U.S.-born as high. What result do you get? (You might also try to work out the calculation taking the cells labeled educated and the cells labeled United States-born as low. It doesn't affect the correlation a bit; statistics don't prove that being educated or being born in the United States is higher than being uneducated or being born in some other country.)

If you remember the work you did before on the relationship between percentages of foreign-born and percentages of uneducated in the nine geographical divisions of the country, you will be surprised at the result you just got. Before, you found a negative correlation; regions with a high percentage of United States-born persons tended to have a low percentage of educated persons. But your new correlation is positive: if an individual in the United States was born in the United States he is more likely to be educated than a foreign-born person in the United States.

Both correlations are correct. The reason why this peculiar reversal has occurred is that the relationship is being affected by something that is not shown in the tables; education and nativity are not related to each other the same way in all parts of the country. Tables 2 and 6 were deliberately included here to warn you against jumping to conclusions on the basis of what the statistics show.[6] Before you draw conclusions, you should try to be sure that a table has not omitted information that must be considered. Can you tell what it is in the present instance? Here are a few hints: the reversal comes about because there is a large group of people in the United States who (a) are native-born, but (b) have not been given as much opportunity to become well-educated as the rest of the population, and (c) are heavily concentrated in a few regions of the country. Can you identify such a group? If so, can you figure out a way of deciding whether it is indeed the source of reversal? State which of the two correlations you think describes the actual situation better.

---

[6] This particular error, of assuming that the correlation between two variables is the same for groups as it is for their individual members, is known as the ecological fallacy. The technical mathematical explanation of how the fallacy arises, using the relationship between nativity and education as an illustration, is given by W. S. Robinson, "Ecological Correlations and the Behavior of Individuals," American Sociological Review, vol. 15 (June, 1950), 351-357.

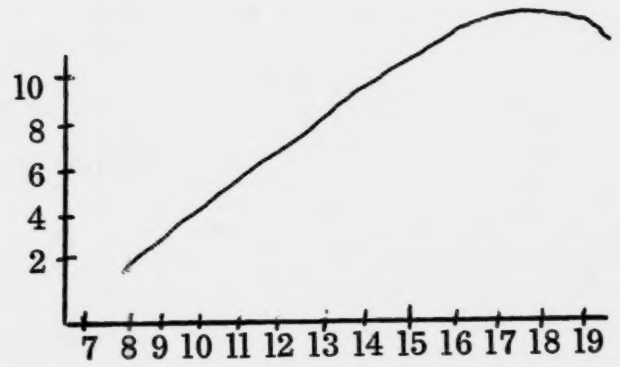
III

SOME ILLUSTRATIVE EXERCISES

A. Explain the following relationship and present your explanation in graphic or tabular form.

B. Design a research project which would test your explanation.

#1 Number of Major Criminal Acts (Self-Reported) Committed During Previous Year

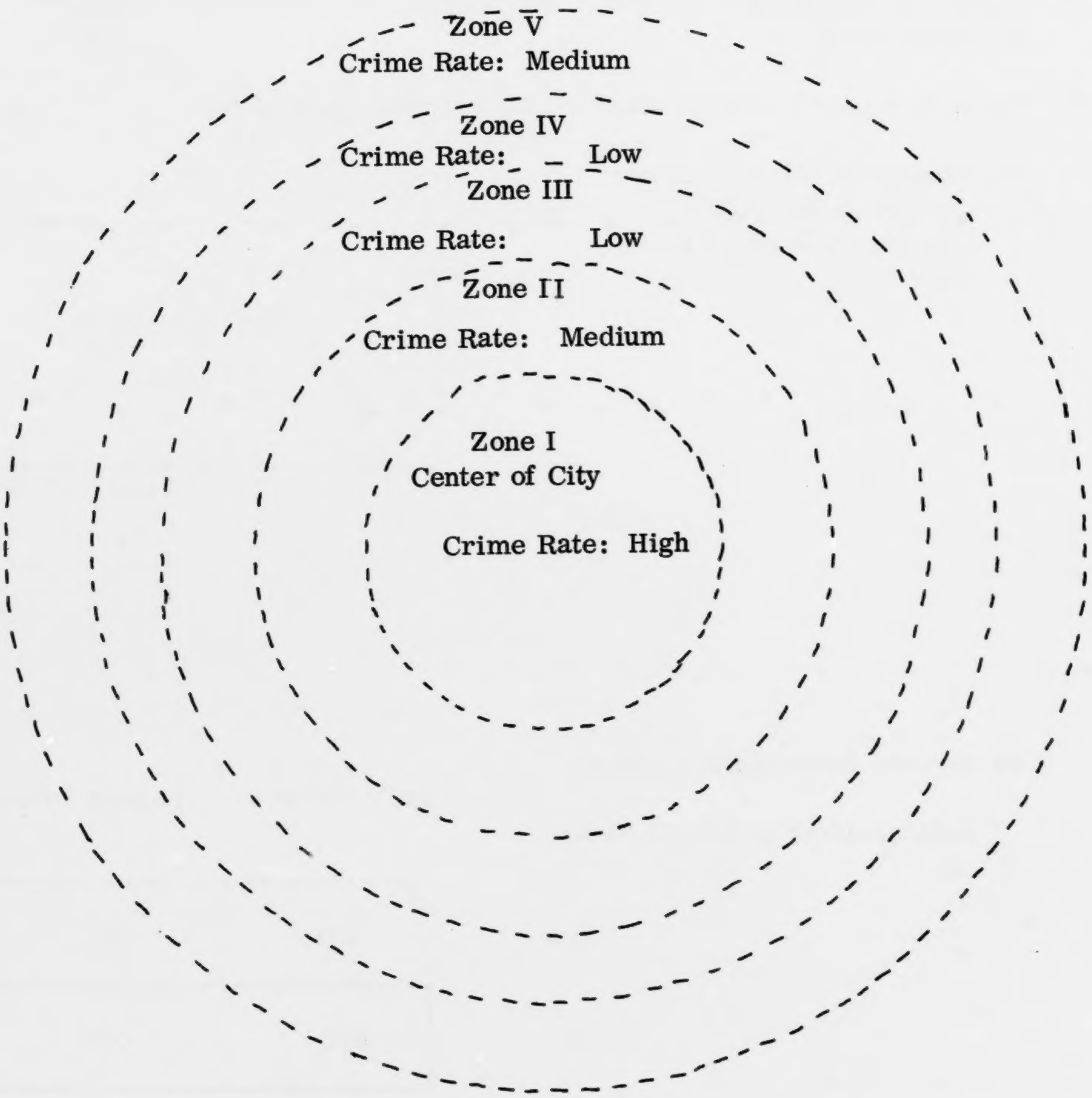


#2 Sex and Automobile Accidents

Auto Accident in Last 12 Months

	Male Drivers	Female Drivers	
Yes	1271	475	1746
No	4152	4496	8648
Total	5423	4971	10394

#3 CITY LOCATION AND CRIME RATE


Zone V
Crime Rate: Medium
Zone IV
Crime Rate: Low
Zone III
Crime Rate: Low
Zone II
Crime Rate: Medium
Zone I
Center of City
Crime Rate: High

Zone I:	Financial, Theatre District, Slum Area
Zone II:	Blighted Area, Small Businesses, Low-Cost Housing
Zone III:	Lower Middle Class Residential Area (Older Buildings)
Zone IV:	Middle Class Residential Area (Newer Buildings)
Zone V:	Suburbs

#4 COMPOSITION OF NEIGHBORHOODS AND ANTI-SEMITISM

Neighborhood	Number of Jewish Families	Number of Non-Jewish Families	Percentage of All Families Expressing Anti-Semitism
Alpha	19	81	16%
Beta	30	71	15%
Gamma	44	58	16%

#5 RANDOM SAMPLES OF SPANISH-SPEAKING AND ANGLO MARRIED COUPLES

	Anglo	Spanish-Speaking
Number of Children Less Than Five Years of Age, Per 1,000 Population	250	424

#6 AVERAGE NUMBER OF VOLUNTARY ASSOCIATIONS OF WHICH THE INDIVIDUAL IS A MEMBER

Socio-Economic Status		
I	(Low)	.9
II		1.2
III		3.5
IV		3.8
V	(High)	2.9

#7

STRICTNESS OF MOTHER'S DISCIPLINE AND SCHOOL DROPOUTS

School Dropout		Strict	Fairly Easy	Easy
	Yes	50	60	75
	No	150	140	125

Same Sample, Broken Down by Sex of Child

		Strict	Fairly Easy	Easy
Boys	Yes	32	32	38
	No	68	68	62
Girls	Yes	18	27	37
	No	82	73	63

#8

YEARLY INCOME AND AGE FOR A RANDOM SAMPLE OF ADULT MALE WAGE EARNERS 25 YEARS OR OLDER, 1960

AGE	MEAN INCOME
25-29	$3200
30-34	$3450
35-39	$3700
40-44	$4200
45-49	$5100
50-54	$5000
55-59	$4700
60 and Over	$2100

#9 YEARLY INCOME AND EDUCATION FOR A RANDOM SAMPLE OF ADULT MALE WAGE-EARNERS, 25 YEARS OR OLDER, 1960

EDUCATION (Years of Schooling)	MEAN INCOME
0-8	$2200
9-11	$3300
12	$3600
13-15	$4500
16	$4600
17 or More	$5200

CHAPTER THREE

THE COLLECTION OF DATA

I

Both law and the social sciences are concerned with the facts of human behavior. The former may be interested in such facts primarily to settle litigation in a specific case and the latter may be interested primarily in using them to formulate scientific generalizations. They unite, however, in that the stuff of which their work is made is not guess or assumption but observed reality.

It is true, of course, that the law and the social sciences have organized their pursuit of "the truth, the whole truth, and nothing but the truth" in quite different ways. Science has usually left the matter to the individual researcher who, in theory, is guided by a disinterested concern for the outcome.[1] The law, on the other hand, has developed a complex social system we call the judiciary with an intricate division of labor and a use of partisan rather than disinterested or neutral truth-seekers. The development of large scale scientific research has changed this picture somewhat, perhaps, but the contrast is still striking.

---

[1] For a discussion of the ethos of science--the institutional norms which govern the search for certified knowledge--see Robert K. Merton, Social Theory and Social Structure (Glencoe, Illinois: The Free Press of Glencoe, 1957), Chapter XVI.

One consequence of the partisan search for the truth in the law is that the accumulation of facts by the court can be hampered as each attorney attempts to get into the record those facts favorable to his side and to suppress those which are not. The attorneys, to be sure, are designated "Officers of the Court," which means that the aim of finding the full truth is supposed to take precedence over the partisan desire to win. We do not have to be overly cynical, however, to suspect that the conflict thus created is often resolved in favor of the attorney's side of the case as suggested by the reading from American Jury Trials.

The court, however, has a variety of devices to further its task; and over the centuries our legal system has hammered out rules of evidence and courtroom procedures which hopefully allow the truth to emerge from the hurly-burly of legal disputes. The prohibition against leading questions and misleading or argumentative questions, as described by Professor McCormick, are examples of how the law attempts to control the collection of data. These procedures of the law, designed to elicit truthful testimony, should be kept in mind by the law student as he engages in the task of carrying out or analysing social science research, for they are a useful reminder of the difficulties which lie in the way of uncovering the facts. But the rules of evidence and the rules of the courtroom cannot be allowed to restrict the demands of research; and the law student must be constantly aware of both the advantages to be gained and the benefits to be lost.

II

There are at least four important ways in which the collection of data in the social sciences differs from the collection of data in the adversary proceedings of the law.

In the first place, "hearsay" testimony is acceptable in a good deal of research and indeed is welcomed--not because it necessarily reflects a true picture of some past event but because it tells us what our respondent believes or feels and these beliefs or feelings, whether erroneous or not, are frequently essential data in uncovering attitudes which may underlie human behavior.

In the second place, the collection of data in the law consists of a very large extent of the collection of oral or written accounts of events. In social science research, however, it is commonly necessary to collect data on behavior first-hand. The problems of inference and proof take on new dimensions as indicated in the article on participant observation by Professor Becker.

In the third place, when the determination of facts takes place in a courtroom (and the collection of facts in litigation is frequently shaped by the expectation of reaching a courtroom even though many cases never get to that stage) testimony can be compelled and subjected to careful cross-examination. In social science research, however, the "collector of facts" stands in a very different relationship to the person who provides the facts. The revelation of attitudes and beliefs is largely voluntary and is heavily dependent on establishing rapport.

In the fourth place, social science research, unlike litigation, is not conducted to win but to fully reveal the truth whatever it might be. The researcher is guided by the norm demanding that he collect and present evidence both favorable and unfavorable to his hypothesis. In a very fundamental sense, his "client" is scientific knowledge and a partisan allegiance to one view or another is forbidden.

The law student, then, is familiar with the problem of collecting accurate information, for it is a problem which he encounters in one form or another in his regular legal studies. But when he moves into the collection of data in social science research, he encounters new problems, new situations, new kinds of facts; and he will be facing intellectual challenges no less great than those encountered in the construction of hypotheses.

III

The variety of issues arising in the collection of data is so great that it is impossible to reduce the process to a neat set of techniques.[2] In the selection from Selltiz, *et al.*, however, the student is alerted to the kinds of problems he will face as well as to some possible solutions. Questionnaires vs. interviews, getting at embarrassing material, the skills of the data collectors, standardized questions, differences in the interpretation of questions due to cultural background or educational level, anonymity, inducements for cooperation, the use of "projective" questions--all of these and more become issues as the researcher develops his plans for collecting data. And it must be admitted that although the social sciences are slowly building a body of research findings on these problems of research, we still lack firm knowledge in this area and many of the problems must be solved in the light of experience and common sense.

One more point is worth noting. Some research is dependent on mailed questionnaires and their use would seem to have an immediate appeal in terms of saving time and money. But mailed questionnaires have their special difficulties, as indicated in the readings from Methods in Social Research by William J. Goode and Paul K. Hatt. Perhaps this will serve to underline the idea that there is no method of collecting data that is without flaws as well as special advantages and that in the choice of methods the researcher cannot fall back on a mindless, automatic procedure, but must constantly bring all of his skills, experience, and imagination into play.

---

[2] Given the limitations of space in this collection of readings, we have been forced to confine ourselves largely to surveys and participant observation as methods of data collection. But there are a number of other methods--perhaps most notably, for the law student, in the use of documents, records, census materials, and so on. See Leon Festinger and Daniel Katz, *Research Methods in the Behavioral Sciences*, New York: Holt, Rinehart, and Winston, Inc., 1953.

INTERVIEWS AND QUESTIONNAIRES

The following reading provides the student with a detailed discussion suggesting when it is best to use the interview as opposed to the questionnaire, how to ask questions aimed at disclosing a respondent's beliefs as opposed to his feelings, and illustrating the various types of interviews and questionnaires being used by social scientists today.

There is one more point which the student should watch for in the reading. Observational methods, it is argued, are less effective in giving information about a person's perceptions, beliefs, feelings, motivations, anticipations, or future plans than questioning. To obtain such information, social scientists have advised the interview schedule, the questionnaire, and a variety of projective methods. According to Professor Selltiz *et al.*, observational methods of empirical investigation are limited by such things as the difficulty of being on hand for relatively rare and spontaneous occurrences and events of long duration. For example, if an anthropologist wishes to learn about a certain kind of religious ceremony by observation rather than by interviews, he may find that his field work is completed long before such an event occurs. Or, as in the case of Skolnick's study, Justice Without Trial, the researcher is limited to observing not the working life of a police officer, but the work of a police officer while the study is in progress. In other words, observational methods force the researcher to rely on an isolated set of events. This is not to say that empirical researchers should not use observational methods but--as we have said before--that a research design should try to make use of more than one means of data collection.

This material has been reproduced from Research Methods in Social Relations by Claire Selltiz, Marie Jahoda, Morton Deutsch, and Stuart W. Cook. Copyright 1951, (c) 1959 by Holt, Rinehart and Winston, Inc. No part of this material may be reproduced in any way without the written permission of Holt, Rinehart and Winston, Inc. pp. 236-268 Reprinted by permission.

Claire Selltiz, et al., Research Methods, pp. 236-268.

> *If we want to know how people feel: what they experience and what they remember, what their emotions and motives are like, and the reasons for acting as they do—why not ask them?*
>
> G. W. ALLPORT

OBSERVATIONAL METHODS, as we have seen in the preceding chapter, are primarily directed toward describing and understanding behavior as it occurs. They are less effective in giving information about a person's perceptions, beliefs, feelings, motivations, anticipations, or future plans; and certainly they provide no information about past behavior or private behaviors, such as sexual activity or dreaming, which are, by their very nature, either unfeasible or impossible to observe. To obtain such information, the interview, the questionnaire, and the projective method have been devised.

In the interview and questionnaire approach, heavy reliance is placed on the subject's *verbal report* for information about the stimuli or experiences to which he is exposed and for knowledge of his behavior; usually the investigator has not observed the events discussed. The subject's report may or may not be taken at face value; it may be interpreted in the light of other knowledge about him or in terms of some psychological theory; inferences may be drawn about aspects of his functioning which he has not reported. Regardless of the amount and kind of interpretation, however, the starting point is the subject's self-report. Thus these approaches can ordinarily obtain only material that the subject is willing and able to report.

For many years now, a controversy has been raging in the psychological literature about the validity of verbal reports. The question is: How do we know a person is *really* hungry when he says, "I am hungry"? There are many weighty issues involved in this question, which we shall not discuss.[1] Let us point out, however, that in everyday life we accept many verbal reports as valid. For example, if we ask a friend what he thought of a certain play and he says "It's terrible," we ordinarily believe that his statement of *feeling* is correct; whether his feeling is appropriate to the play is, of course, another matter. However, in everyday life we also realize that in certain circumstances verbal reports are not to be trusted. For example, whenever we have reason to suspect that a person's truthful self-report would be embarrassing, humiliating, or degrading, or would in some way place him in an unfavorable light, we are likely to entertain some reservations about a report that shows him in favorable light. Or whenever we have reason to believe that a person is using a verbal report to ingratiate himself, to gain respect or prestige, to amuse or astonish, or in some other way to create a certain social effect, we are likely to place little confidence in it. In other words, when the circumstances in which the report occurs lead us to suppose that the subject's *motivation* or the pressures to which he is exposed are such as to prevent a candid report, we are not likely to give it much credence.[2]

Not only may people be reluctant to report openly their beliefs, feelings, motivations, plans, and so on; they may be *unable* to do so. As psychoanalysts have pointed out, we are not aware of many of our important beliefs and motivations, and hence cannot report them. Moreover, self-report frequently requires *self-diagnosis*. Even such seemingly simple questions as "Are you shy with strangers?," or "Would you rather go to a party or stay home and read a good book?" require the individual to make a judgment about himself on the basis of many past events. Feelings, beliefs, and motivations become apparent to the self in an intellectually comprehensible form only as the end result of an involved process of inference. With respect to complex social attitudes, many people have never learned to make the inferences necessary to an adequate verbal report; they cannot indicate, in any systematic or analytic manner, their attitude toward their husband or wife, for example, or toward "progressive education," or toward a minority group.

---

[1] For an excellent discussion of these issues, see the "Symposium on Operationism" in the *Psychological Review*, 1945, 52, 241–294.

[2] The research of Parry and Crossley demonstrates how the answers even to factual questions may be influenced by the desire to appear "respectable." As summarized by Katz (1951), "They found that people consistently exaggerated their registration and voting behavior. The exaggeration varied from thirteen percent who falsely claimed to have voted in the 1948 election to twenty-eight percent who made fictitious claims to voting in local elections. One third of those who reported contributing to the Community Chest were speaking of pious intentions, not actual contributions. Telephone and home ownership were accurately reported. Similarly, car ownership was not appreciably inflated, but ten percent of those reporting a driver's license did not have one. Again, ten percent claimed to have library cards when in fact they had none."

Nevertheless, every person has a unique opportunity to observe himself. To the extent that he can and will communicate his knowledge about himself, he provides the investigator with information that could otherwise be obtained, if at all, only by more time-consuming methods. Despite the limitations of self-report, it is frequently both possible and useful to get an individual's own account of his feelings toward a psychological object, his image of the object, his views of appropriate behavior toward it, etc.

Comparison of Interview and Questionnaire

Although both interviews and questionnaires place heavy reliance upon the validity of verbal reports, there are important differences between the two methods. In a questionnaire, the information one obtains is limited to the written responses of subjects to prearranged questions. In an interview, since the interviewer and the person interviewed are both present as the questions are asked and answered, there is opportunity for greater flexibility in eliciting information; in addition, the interviewer has the opportunity to observe both the subject and the total situation to which he is responding. Let us detail further some of the general characteristics of the questionnaire and interview approaches, with their respective advantages and disadvantages.

ADVANTAGES OF QUESTIONNAIRES

By its very nature, the questionnaire is likely to be a less expensive procedure than the interview. It requires much less skill to administer than an interview; in fact, questionnaires are often simply mailed or handed to respondents with a minimum of explanation. Further, questionnaires can often be administered to large numbers of individuals simultaneously; an interview, on the other hand, usually calls for questioning each individual separately.[3] Questionnaires can be sent through the mail; interviewers cannot.[4] With a given amount of funds, it is usually possible to cover a wider area and to obtain information from more people by means of questionnaires than by personally interviewing each respondent.

The impersonal nature of a questionnaire—its standardized wording, its standardized order of questions, its standardized instructions for recording responses—ensures some uniformity from one measurement situation to another.[5] From a psychological point of view, however, this uniformity may be more apparent than real; a question with standard wording may have diverse meanings for different people, may be comprehensible to some and incomprehensible to others. Nevertheless, much can be done to ensure meaningful uniformity of questions by careful pretesting and by helping the subjects to understand the questionnaire during its administration. The interviewing situation, on the other hand, is rarely uniform from one interview to the next. Not only do the personalities of different interviewers affect the measurement situation differently; each interviewer is bound to vary somewhat from interview to interview. Moreover, in some types of interview, the

[3] This is not true, of course, of group interviews, where as many as eight to ten people may discuss the subject matter of an investigation under the direction of an interviewer. However, such interviews are more satisfactory as a source of hypotheses or as a way of gathering information about the *group;* they do not ordinarily yield systematic information from every individual in the group on each point covered in the interview schedule.

[4] Interviewing by telephone may, in certain circumstances, cost less per return than the mail questionnaire. The telephone interview is particularly useful in obtaining information about what an individual or a family is doing (e.g., what television program he is watching) at the time of the call. Usually, telephone interviewing has to be brief and superficial to obtain the cooperation of the respondent. Another serious limitation is that telephone surveys cannot reach a random sample of the total population, since not all people have telephones, and people who work away from home are hard to reach by telephone. For a further discussion of telephone surveys see Blankenship (1946) and Parten (1950).

[5] This is not true for mailed questionnaires, where there is likely to be considerable variation from home to home in the conditions under which the questionnaire is filled out. In one home, for example, the questionnaire may be filled out by the head of the family, in another by some other member; in one the questionnaire may be given time and attention, in another it may be competing with a television broadcast or a crying baby.

interviewer has no standard set of questions to ask. As a result, interviews may be less easily comparable with one another than questionnaires.

Another advantage of questionnaires is that respondents may have greater confidence in their anonymity, and thus feel freer to express views they fear might be disapproved of or might get them into trouble. Although an interviewer may assure the respondent that he will not be identified in any way, the respondent may doubt his good faith; since, in most interviewing situations, the interviewer knows either the respondent's name or his address or other identifying information, it is always possible that he may include this information in the completed interview. If a questionnaire is presented as anonymous and there is no apparent identifying information, the respondent may feel greater confidence that his replies will not (or cannot) be identified as coming from him. Studies that have used both methods have sometimes found rather marked differences between the replies to the interview and those to the questionnaire. Edwards (1957a), for example, in a survey of attitudes of residents of Seattle toward a proposed state bill providing a cash bonus to war veterans, had interviewers question half of the respondents; the other half were given a sheet marked "Secret Ballot," which they checked, folded, and inserted into a box labeled "Secret Ballot Box." (This was, of course, an unusually vivid way of emphasizing anonymity.) The interviews showed many more "don't know" replies and fewer unfavorable responses than the ballots. A referendum vote in an election held a few weeks later provided a check; the proportions obtained from the secret ballot were much closer to those of the actual vote than the results of the direct interview. It should be noted, however, that anonymity is not always the best method of inducing frank answers. On more complex questions, where there may be strong emotional involvement—as, for example, questions of marital adjustment—an understanding and permissive manner on the part of an interviewer is likely to be more successful than the anonymity of a questionnaire in eliciting frank responses.

Another characteristic of the questionnaire that is sometimes, though not always, desirable is that it may place less pressure on the subject for immediate response. When the subject is given ample time for filling out the questionnaire, he can consider each point carefully rather than replying with the first thought that comes to mind, as often happens under the social pressure of long silences in an interview.

ADVANTAGES OF INTERVIEWS

It has been estimated that, for purposes of filling out even simple written questionnaires, at least 10 per cent of the adult population of the United States is illiterate. For complex questionnaires, the percentage would undoubtedly be considerably higher.[6] Thus, one of the major drawbacks of the usual questionnaire is that it is appropriate only for subjects with a considerable amount of education. Complicated questionnaires requiring extended written responses can be used with only a very small percentage of the population. Even many college graduates have little facility for writing, and of those who do, few have the patience or motivation to write as fully as they might speak. Hence, questionnaires are not an appropriate method for large segments of the population; for those for whom they are appropriate, the burden of writing or of maintaining interest is great enough to limit the number of questions that may be asked and the fullness of the responses. On the other hand, interviews can be used with almost all segments of the population; in fact, in contrast with the questionnaire, a frequent problem in interviewing is that of limiting the responses of the verbose individual.

Surveys conducted by personal interviews have an additional advantage over surveys conducted by mailed questionnaires in that they usually yield a much better sample of the general population. Many people are willing and able to cooperate in a study when all they have to do is talk. When questionnaires are mailed to a random sample of the population, the proportion of returns is usually low, varying from about 10 to 50 per cent. There are many factors that influence the percentage of returns to a mailed questionnaire. Among the most important are: (1) the sponsorship of the questionnaire; (2) the attractiveness of the questionnaire format; (3) the length of the questionnaire; (4) the nature of the accompanying letter requesting cooperation; (5) the ease of filling out the questionnaire and mailing it back; (6) the inducements offered to reply; (7) the nature of the people to whom the questionnaire is sent. Attractively designed questionnaires that are short, easy to fill out, simple to return, sponsored by a group with prestige, and presented in a context that motivates the respondent to cooperate are most likely to be returned. However, even under the best of circumstances a sizable proportion do not return questionnaires. The people who do return them are usually the less mobile (and thus the more likely actually to receive the questionnaire), the more interested, the more literate, and the more partisan section of the population.[7]

---

[6] Subjects with limited education may be able to fill out questionnaires with the help of questionnaire administrators. However, in such cases the questionnaire loses much of its advantage over the interview with respect to economy.

Another advantage of the interview is its greater flexibility. In a questionnaire, if the subject misinterprets a question or records his responses in a baffling manner, there is usually little that can be done to remedy the situation. In an interview there is the possibility of repeating or rephrasing questions to make sure that they are understood or of asking further questions in order to clarify the meaning of a response. Its flexibility makes the interview a far superior technique for the exploration of areas where there is little basis for knowing either what questions to ask or how to formulate them.

In addition, the interviewing situation offers a better opportunity than the questionnaire to appraise validity of reports. The interviewer is in a position to observe not only *what* the respondent says but also *how* he says it. He can, if he wishes, follow up contradictory statements. If need be, the interviewer can directly challenge the subject's report in order to see how consistent his answers will be.

The interview is the more appropriate technique for revealing information about complex, emotionally laden subjects or for probing the sentiments that may underlie an expressed opinion. If a verbal report is to be accepted at face value, it must be elicited in circumstances that encourage the greatest possible freedom and honesty of expression. Although, as already noted, an anonymous questionnaire may sometimes be the most effective way of producing such a permissive atmosphere, its usefulness is limited to issues on which respondents have rather clearly formulated views that can be simply expressed. The more or less rigid structure of questionnaires, the inability to explain fully in writing one's asocial or antisocial feelings and behavior, and the solemnity and permanent nature of a response that is put on paper in one's own handwriting or (if the questionnaire is not anonymous) under one's own name—all work against frank discussions of socially taboo or socially controversial issues in response to a questionnaire. With respect to many questions, an interview is likely to be more successful in creating an atmosphere that allows the respondent to express feelings or to report behaviors that are customarily disapproved.[8]

---

[7] For a fuller discussion, see Parten (1950, Chapter 11).

In the interview situation, the "social atmosphere" can be varied in other ways. Behavior in real life occurs in situations that are seldom free from social pressures. The interview, more than the questionnaire, allows one to approximate in the measurement situation these varying social pressures, since the interviewer can, within limits, vary the nature of the atmosphere as he questions the respondent. He can, for example, point out objections to the position of the person being interviewed, and observe how the latter responds.[9] This is a very useful flexibility, especially if the ultimate objective of the measurement is to predict behavior in varied situations.

Question Content

In both questionnaires and interviews, information is obtained by asking questions. Questioning is particularly suited to obtaining information about what a person *knows, believes* or *expects, feels* or *wants, intends* or *does* or *has done*, and about his *explanations* or *reasons* for any of the preceding. These major types of question content are discussed below. It should be recognized, however, that questions do not always fall neatly into one or another "content type." The distinctions among types are a matter of custom and convenience rather than theoretical rigor.

---

[8] For an interesting discussion of the flexible use of the interview method to obtain information about a socially taboo subject, see Kinsey *et al.* (1948).

[9] As a rule, this would constitute very bad interviewing practice. However, in interviews designed to measure responses under varying degrees of social pressure, such procedures on the part of the interviewer are an essential part of the measurement process.

CONTENT AIMED MAINLY AT ASCERTAINING "FACTS"

Often the simplest and most economical method of obtaining "facts" is to go directly to the people who are in a position to know them and to ask for the desired information. It is reasonable to assume that people who have access to information, who are sufficiently intelligent to absorb it, and who are motivated to acquire and retain it are able, if they are willing, to provide the investigator with reports of many interesting and valuable "facts." We may expect, for example, that the people who are responsible for the execution of a policy know what it is. Similarly, we should expect every individual to know a variety of facts about himself and his environment. A sizeable proportion of most questionnaires and interviews is directed toward obtaining such facts. Questions about the person's age, education, religion, income, nationality, marital status, occupation, etc., are of this type. So, too, are questions about the characteristics (behavior, beliefs, feelings, desires, intentions, etc.) of persons who are known to the respondent, such as family, friends, and colleagues. Questions about events, circumstances, policies, etc., known to the respondent are also of this nature.

Reported "facts" must, of course, always be evaluated in terms of credibility. The rules of evidence that have developed through the centuries in judicial procedure are a good source of insight into factors affecting credibility. Thus, it is always pertinent to raise such questions as: How did the respondent obtain knowledge of the "fact"—through direct observation, through inference, through hearsay, etc.? What motives may the respondent have in reporting the "fact"? How accurate is the respondent's memory of the "fact" likely to be? Any one of these factors may affect, for example, even such an apparently clear-cut piece of information as a housewife's report of her husband's income. Her knowledge may be based only on her husband's statement, and he may be deceiving her; or she may wish to impress the investigator and therefore claim a higher income; or she may not be interested enough in her husband's income to remember it precisely. The fallibility of memory for nonrecurring events, for events in the distant past, for events of little interest, and for events difficult to comprehend; the ephemeral quality of memory and its dependence on situational factors; the corruptibility of memory in relation to events of significance to the self—all of these factors require caution in accepting as true the remembrance of things past

When the focus is on description or understanding of an event, a situation, or a community, rather than on information about the individual respondent, it is frequently possible to check accuracy through comparison of the reports of several respondents. If respondents occupying widely different positions in the community agree on a statement, there is much better ground for accepting it as true than if only one of these respondents makes the statement. On the other hand, contradictions between the reports of apparently reliable informants provide important leads for further investigation. Rapkin, Grier, and Grier (1957) used this method of checking in a survey of race relations in an urban community. The study was exploratory in nature, the purpose being to identify problems on which research was needed as a basis for policy decisions. The investigators interviewed a variety of individuals in the community: city officials, representatives of Negro organizations, white persons known to be concerned with problems of interracial relations, other community leaders. The respondents were asked about such matters as the extent of employment of Negroes, police treatment of minority groups, etc. On some points the reports were almost unanimous; on others there was wide discrepancy. Almost all informants, for example, agreed in their accounts of the policy and practices of the Public Housing Authority with respect to racial integration, though there was some disagreement about the wisdom of that policy. This impressive agreement led the investigators to conclude that there was no need to study further the nature of the Housing Authority's policy and practices, though there might well be need for research directed toward finding ways of solving the problems the agency was encountering in carrying out its policy. At the other extreme, reports about police treatment of Negroes were contradictory; they ranged from the statement that Negroes were arrested for even the most minor infractions, which would go unnoticed if committed by whites, to the view that law enforcement in Negro neighborhoods was so lax that only the most serious offenses brought police action. On the basis of this striking lack of agreement about the actual state of affairs, investigation of the practices of the

Police Department with respect to Negroes was included in the recommendation of research priorities.

Often, however, contradictions can be cleared up within a given study. Whenever possible, statements should be checked against statistical records. Comparison of contradictory statements may provide a clue to the accuracy of one of them. For example, Rapkin and his associates were told by a representative of a minority group organization that the local breweries did not employ Negroes. Checking with the appropriate union brought the explanation that no new brewery personnel of any kind had been hired within the past ten or twelve years; employment had been at a peak during World War II, and when additional workers were needed in the subsequent period, former employees were called back. Thus, if there was discrimination against Negroes, it was a reflection of a situation that had existed some years before, not necessarily of present policies. Several other informants supported the union's version; the person who had originally made the statement about discrimination, when presented with the union's explanation, agreed that it was correct.

Content Aimed Mainly at Ascertaining Beliefs about What the Facts Are

Instead of asking questions to find out the objective facts from people in a position to know, the investigator may wish to learn what people believe to be the "facts." This is the purpose, for example, in asking a respondent to indicate whether the following statement is true or false: "No Negro has ever made a worth-while contribution to the arts." The respondent's answer is not used to establish what is objectively true but rather to provide a picture of his beliefs. Frequently, before asking questions about the *nature* of a person's beliefs, it is desirable to find out whether he *has* any beliefs or information relevant to the topic under investigation. For example, it makes little sense to inquire, except "projectively," about a person's beliefs about the United Nations if he does not know that the organization exists.

The distinction between an inquiry into "facts" and an inquiry into beliefs must be kept clearly in mind by the investigator. If, for example, he wishes to know objectively how much delinquency there is in a community, he will interview people who have been carefully selected for their knowledge about the topic—members of the police force, for example, workers in settlement houses or recreation centers, group workers assigned to street gangs (if there is such a program in the community)—and will consult court, police, and school records, etc. On the other hand, if he wishes to know what people *believe* about the extent of delinquency in the community, he will interview a general sample of the population rather than selected "experts," and he may simply ask about their beliefs rather than pressing for objective evidence. To use a simple analogy, one does not measure the temperature of a room by asking the people in it how hot they believe it is. But if one is interested in the subjective experience of temperature under varying conditions, one may follow precisely this procedure of asking people how hot they believe the room is. In the field of social attitudes, the relationship between objective reality and a person's beliefs is frequently of considerable interest. Distortions in perception and beliefs, as well as gaps in knowledge, are very often clues to a person's desires or fears.[10]

In addition to discovering the content of a person's beliefs, a questionnaire or interview may provide information about the characteristics and interrelationships or structure of the beliefs. In other words, the purpose of questioning may be to investigate such aspects of belief as: What are the person's most important beliefs? Which beliefs are relatively private, and which are accessible to the public? How specific is a given belief? How clear is it? How strongly held?[11]

CONTENT AIMED MAINLY AT ASCERTAINING FEELINGS

A person's beliefs about what the facts are will often give very clear indications of his feelings and his desires. The converse is also true; an emotional reaction will sometimes reveal beliefs that a subject is unable to verbalize. To understand a person's behavior, knowledge of his feelings may be at least as fruitful as knowledge of his beliefs.

In questionnaires, perhaps the most common method of investigating feelings is to include items that bear directly on various possible emotional reactions—fear, distrust, disgust, contempt, hate, envy, sympathy, admiration, etc. Some examples follow:[12]

---

[10] For a discussion of techniques that employ distortions in perception and memory as a method of measuring social attitudes, see Chapter 8.

[11] For a discussion of the measurement of the various characteristics of beliefs, see Krech and Crutchfield (1948).

When prize fights are held between Negroes and white men, I want the Negro to win.

The sight of a Negro almost always frightens me.

Feelings and motives are probably investigated better, however, by questions that allow the subject considerable freedom in response. Emotional reactions are frequently too complex to report in a single phrase. Moreover, the words used to identify an emotional reaction may not have the same meaning for the investigator and for the respondent.

Questions that call for simple unitary responses assume that a subject reacts with the same emotion toward members of a given group regardless of the specific situation or of his relationship with specific members of the group. Clearly, this is a dubious assumption. For example, a highly prejudiced plantation owner in the United States is likely to feel somewhat benign toward Negroes who are self-abasing and "Uncle Tom-ish" but hostile toward Negroes who refuse to consider themselves his inferiors. An investigation of emotional reactions, if it is to provide a full picture, must uncover not only the individual's feelings but also the circumstances in which the feelings are likely to be aroused. Both can be studied most concretely by linking them to specific events in the subject's past. Thus, instead of asking, "How do you feel about walking through Negro neighborhoods alone at night?," one would ask, "Did you ever walk through a Negro neighborhood alone at night? (If yes) How did you feel?"[13] Of course, if the investigator is interested in the subject's response to an abstract "Negro neighborhood," the former phrasing may be preferable.

CONTENT AIMED MAINLY AT DISCOVERING STANDARDS OF ACTION

An individual's definitions of appropriate behavior in various social situations are of interest both as a reflection of the prevailing your establishment?" Responses were received from half of the establishments; of these, over 90 per cent said "No."

---

[12] For further illustrations, see Kramer (1949).
[13] For further illustrations, see Kramer (1949).

On the other hand, Pace (1939) found a high correspondence between answers to a series of "what would you do" questions on social, economic, and political issues, and actual behavior as indicated by voting and group membership. He presented college students with 37 questions, of which an example is given here:

> Your state needs an additional tax so that its budget can be balanced. If you could vote, and a tax bill which provided for a widely distributed general retail sales tax was submitted to the people of the State for approval, what would you do?
> _____not vote at all
> _____vote for the bill
> _____vote against the bill
> _____vote for the bill and try to persuade others to vote for it too
> _____vote against the bill and try to persuade others to vote against it too

In a pretest, Pace administered this questionnaire to 25 known "radicals" and 25 known "conservatives." The "radicals" included members of the Young Communist League, Trotskyites, Farmer-Laborites, liberal New Dealers; they had all been seen frequently at partisan meetings and at addresses by such speakers as Earl Browder and Norman Thomas; they had all voted Communist, Socialist, or Farmer-Labor in 1936. None of the "conservatives" had ever been seen by the investigator at such meetings; 23 of them had voted for Landon in 1936. Scores on the questionnaire distinguished between the two groups with no overlap. Moreover, on all but three of the 37 items, the mean scores of the two groups differed significantly.

Interesting differences have been found between answers to "should" and "would" questions. In a study by Blankenship (1946), at a time when an amendment to permit horse racing and parimutuel betting was under consideration, alternative question wordings were used with two equivalent samples. One set of respondents was asked: "Is it desirable to permit or to prohibit horse racing and parimutuel betting in New Jersey?" The other was asked: "Would you vote for or against the amendment to permit horse racing and parimutuel betting in New Jersey?" The first received more "don't know" and more opposed answers. This suggests that the "is it desirable" form was answered from a social or moral point of view, whereas the "would you vote" form was answered in terms of personal preference.

In a study of reactions to prejudiced remarks (Selltiz *et al.*, 1950), subjects were shown a skit representing an informal situation in which an anti-semitic remark was made before a group of people. In interviews following the skit, the subjects were asked a series of three questions: "What do you think is the right thing to do or say? What do you think you yourself really would have done in a situation like this? What do you think most people would do or say in a situation like this?" More than half (56 per cent) of the respondents replied that the *right* thing would be to answer the anti-semitic remark in some way (that is, verbally express disagreement with it); only 35 per cent said *they themselves* would have answered the remark; and only 15 per cent said *most people* would answer the remark.[14]

CONTENT AIMED MAINLY AT PRESENT OR PAST BEHAVIOR

The present or past behavior of any person is a type of "fact" that he himself is in a uniquely favorable position to observe. We single out this type of "fact" for special notice because of the value of knowledge of past and present behavior in predicting future behavior. How a person has behaved in the past in a certain type of situation is, in the absence of contradictory evidence, an indication of what his future behavior will be in similar situations.

In asking about present or past behavior, experience has demonstrated that the most valid answers are obtained by specific rather than general questions. For example, it is preferable, in a study of consumer behavior, to ask, "Which brand(s) of coffee do you have in the house at the moment? May I see it? Do you usually buy this brand?" than to ask, "Which brand of coffee do you usually use?" Specifying a concrete instance and then asking whether this instance is typical or atypical provides the subject with more cues for recall climate of opinion and as a basis for predicting his probable behavior in such situations. Definitions of appropriate action frequently have two components: ethical standards of what *should* be done, and practical considerations of what it is *feasible* to do. Questions may be directed toward either of these components.

---

[14] Footnote omitted.

The following are examples of questions focused on what *should* be done (Kramer, 1949):

> For years ____________ College, a privately endowed college in New England, had followed the policy of admitting Jewish students in numbers no greater than eight percent of the Freshman class. When a new president took office he dropped this policy and proposed to admit Jews on the same basis as other students. Do you approve or disapprove of the action of the new president? Why?
>
> In your opinion should Negro and white children go to the same public schools, or to separate ones? Why do you think this?

The "should" or "ought to" question—whether phrased in terms of "I," "he," "we," or "they"—provides an indication of the idealized policies of the individual, of the actions he would favor in a situation free from all but moral imperatives. These policies are, for the most part, the product of the idealistic social expectations to which the individual has been exposed in the teachings of societal surrogates—parents, clergy, teachers, government spokesmen. When there is a large discrepancy between social ideals and practices, the ideals, for most adults, tend to function as a guide to what to say on formal occasions rather than what to *do* in everyday behavior (see Lee, 1949); they may also be the source of "twinges of conscience" with respect to everyday behavior.

A person's behavior toward any person or group is determined not only by his beliefs, feelings, and social conscience but also by what he considers to be feasible behavior in the existing social situation. Thus, in parts of the United States or of South Africa, an unprejudiced person in the face of existing social pressures against being friendly with Negroes may act in a manner resembling that of a prejudiced person. Conformity needs, as well as beliefs and feelings about the individual or group that is the ostensible "object" of the behavior, are likely to be important determinants of action. Thus it is useful to inquire not only into beliefs and feelings about what is "right" but also into the "realistic" policies that serve to guide the individual's actions in specific situations.

For example, Stouffer (1949), in a study of conflicting social norms, asked students to imagine that they were proctoring an examination and saw a fellow student cheating. A questionnaire listed possible steps the proctor might take, ranging from "Take away his notes and exam book, dismiss him and report him for cheating," to "Act as if nothing had happened and *not* report him for cheating." Each student was asked to indicate which of these actions he, as proctor, would be most likely to take, next most likely, etc., under a number of specific conditions: If he did not know the student who was cheating; if the student were a good friend of his; if the authorities but not his student friends were likely to hear about his action; if his student friends but not the authorities were likely to hear about his action.

The "would" question involves a personal prediction with respect to one's behavior in a given situation. Unless the respondent has been in a comparable situation at some time, he may have little basis for making such a prediction. His response in this case may be no more than an expression of his desires or of his moral standards. Behavior in a real-life situation is influenced by many momentary social pressures that are difficult to imagine unless they have been previously experienced. Personal prediction by a respondent may thus be a hazardous affair unless it is solidly based on past behavior.

As a matter of fact, even when a respondent *has* in the past been faced with a situation of the type described, his reply about what he would do may be at variance with his actual past behavior or his probable future behavior. The classic investigation of the discrepancy between actual behavior and response to a question of the "what would you do" type is that of LaPiere (1934). LaPiere traveled throughout the United States with a young Chinese couple, stopping at 66 hotels, auto camps, and tourist homes, and eating in 184 restaurants and cafes. Only once were they refused service. Six months later, LaPiere sent a questionnaire to all of the places at which they had eaten or slept, asking, "Will you accept members of the Chinese race as guests in and, in a sense, binds him to a reality that acts as an obstacle to distortion in response. Similarly, in the study of prejudice, questions about past behavior in specific situations are likely to elicit a more accurate report than general inquiries about previous behavior. Thus it is preferable to ask, "For whom did you vote in the last mayoralty election? What made you vote for him? Did you know the religion of the candidates? Were you influenced for or against any candidate by knowledge of his religion? Do you usually tend to consider a candidate's religion in deciding for whom to vote?" than to ask only the general question, "Do you usually tend to consider a candidate's religion in deciding for whom to vote?"

CONTENT AIMED MAINLY AT CONSCIOUS REASONS FOR BELIEFS, FEELINGS, POLICIES, OR BEHAVIOR

Finally, questions may be designed to obtain the reasons the respondent is able to offer for his beliefs, feelings, policies, or behavior. In effect, the investigator is interested in finding out *why*. "Why?" may seem like a simple question, but, as Lazarsfeld (1935) has pointed out, the answer to it is seldom simple. Consider merely the problem of determining why one student selects one college, another a different one. A full answer to the question would require knowledge of their information about various colleges, of their needs and interests, of their ability to meet the entrance requirements or the financial cost of different colleges, and perhaps of still other factors.

In order to secure a full answer to the question "Why?" it is well for the investigator to consider the various possible factors that may influence the belief, behavior, etc., in which he is interested, and to provide for consideration of each factor by asking a number of specific questions rather than a single "Why?" Although the specific influences that are relevant depend on the question being studied, it is possible to identify certain broad classes of considerations that are likely to enter into the determination of *why*. They include: (1) The *history* of the act or feeling: e.g., "What were the circumstances when you started to . . .?" (2) The *characteristics in a given entity* that provoke a given reaction: e.g., "What is there about______that leads you to (feel, believe, act, etc., in a given way)?" (3) The *supports* for the beliefs, feelings, etc., about a given entity: e.g., "What do your friends, relatives, clubs, etc. (feel, believe, etc.) about ______ ?" "What evidence is there to support your beliefs, feelings, etc., about______?" (4) The *personal desires, motives, values, or interests* involved in a given reaction: e.g., "Is there anything about yourself that makes you want to (believe, feel, or act in a given way)?" (5) The *specific situations and circumstances* in which a given reaction occurs: e.g., "In what types of situation are you most likely to (feel or act in a given way)?"

In addition to these reasons *for* a given belief, feeling, action, etc., it may be relevant to inquire into the reasons *against* alternative beliefs, actions, etc. It may also be important to distinguish between past and present influences; for example, between reasons for starting on a given course of behavior and reasons for continuing it.

Once the investigator has decided which kinds of influences are likely to be relevant to his particular question, he sets up an "accounting scheme" (see Zeisel, 1957), mapping out, in preliminary fashion, the various kinds of reasons in which he is interested and providing questions to tap each of them. The following illustration of a set of questions to serve as a guide in an interview aimed at learning why an individual selected a particular college is adapted from Zeisel. Note that it starts with the general question, "Why?," permitting the investigator to find out what is salient in the mind of the respondent, and then provides specific questions to cover the history of the choice and the kinds of influences in which the investigator is interested but which the respondent may not have mentioned in his spontaneous reply:

1. Why, when planning your college years, did you decide to go to______ college?
2. (Supplementary questions clarifying the history of the decision. Ask only those that were not answered in 1.)
   a. When did you first seriously consider going to ______ college?
      (Probe for answer in terms of circumstances rather than dates.)
   b. Did you seriously consider any other colleges you might go to?
   c. How did you come to your decision?

3. (Supplementary questions about reasons for the choice. Ask only those that were not answered in 1 or 2.)
   a. Was there anything about the college itself (and/or the alternatives) that influenced your decision? What? How did you learn about these particular qualities of the college?
   b. Did you feel that any of your own particular needs would be particularly well taken care of by this choice? (Needs while at college? Postgraduate needs?)
   c. Did your parents, friends, teachers or any other persons help you to come to your decision? In what way? Did your own decision in any way depend on the decisions made by your friends? Did you work it out together?
4. In the light of your actual experience would you now make a different choice? Would you make the same choice on a different basis?

Types of Interviews and Questionnaires

The *form* of interviews and questionnaires may vary widely. Interviews may range all the way from the rigidly *standardized*, in which both the questions and the alternative responses permitted the subject are predetermined, to the completely *unstructured*, in which neither the questions to be asked nor the responses permitted the subject are determined before the interview. Although the possible range of questionnaire structure is more limited, there too some variation is possible.

STANDARDIZED INTERVIEWS AND QUESTIONNAIRES

In the standardized interview or questionnaire, questions are presented with exactly the same wording, and in the same order, to all respondents. The reason for standardization, of course, is to ensure that all respondents are replying to the same question. If one interviewer asks, "Would you like to see taxes reduced next year?" and another asks, "Do you think a tax reduction next year would be desirable?," the answers may not be comparable. Differences in question order can also influence the meaning and implications of a given question. The question about the desirability of a tax cut might well be answered differently if it followed a question about the need for developing intercontinental missiles than if it followed a question about the respondent's budget.

Standardized interviews and questionnaires may differ, however, in the amount of structuring of the questions used. They may present fixed alternative answers, or they may leave the respondent free to answer in his own words.

"FIXED-ALTERNATIVE" QUESTIONS. A "fixed-alternative" (or "closed") question is one in which the responses of the subject are limited to stated alternatives. These alternatives may be simply Yes or No, or they may provide for indicating various degrees of approval or agreement, or they may consist of a series of replies of which the respondent picks one as being closest to his position. The following are examples of fixed-alternative questions:

> To what social class would you say you belong—middle class, lower class, working class, or upper class?
>
> For the purpose of our survey, we need to have a rough indication of the income of your family. Would you mind telling me in which of these classes it falls:
>
> Below $1,000 a year
> From $1,000 to $2,000
> From $2,000 to $3,000
> From $3,000 to $4,000
> From $4,000 to $5,000
> Over $5,000 a year
>
> Put a *1* in front of the thing that is most important to have or do in order to get ahead in the world. Put a *2* before the next most important, etc.
>
> ____ pull
> ____ good luck
> ____ brains
> ____ hard work
>
> Do you own an automobile? ____ Yes ____ No
>
> As you probably know, the Governor of Arkansas has called out the National Guard to prevent Negro children from attending a formerly all-white high school in Little Rock. How do you feel about this action? Would you say you:
>
> ____ strongly approve
> ____ mildly approve
> ____ are undecided
> ____ mildly disapprove
> ____ strongly disapprove

Questions of this type are essentially the same whether they are used in interviews or in questionnaires. The only reason for using interviews rather than written questionnaires with this type of material is to reach subjects who are either not willing or not able to fill out questionnaires.

"OPEN-ENDED" QUESTIONS. The "open-ended" question is designed to permit a free response from the subject rather than one limited to stated alternatives. The distinguishing characteristic of open-ended questions is that they merely raise an issue but do not provide or suggest any structure for the respondent's reply; the respondent is given the opportunity to answer in his own terms and in his own frame of reference.

Examples of open-ended interview questions follow:

Now that you have been living in ______ for ______ years,
I wonder if you'll tell me how you feel about it?
  a. What do you like *most* about it?
  b. What do you *dislike* most about it?
  c. How about the neighborhood? What do you think of it?
  d. How about living in a public housing project?

When used in questionnaires, both the questions and the order in which they are presented are predetermined; it is impossible to ask any supplementary questions. When open-ended questions are used in standardized interviews, the questions and their order are predetermined, but the interviewer is given freedom to repeat the question if the reply is not to the point and to use at his discretion such nondirective probes as, "Won't you tell me more? What makes you think . . .? Why? In what way . . .?," etc. The task of the interviewer is to encourage the respondent to talk freely and fully in response to the questions included in the interview schedule and to make a verbatim record of his replies. Generally he has no freedom to raise new questions except to clarify the meaning of the subject's responses, and these must be nondirective.

ADVANTAGES AND DISADVANTAGES OF OPEN-ENDED AND FIXED-ALTERNATIVE QUESTIONS. Fixed-alternative questions have the advantages of being "standardizable," simple to administer, quick and relatively inexpensive to analyze. The analysis of responses to open-ended questions is often difficult and expensive. Categories for analysis must be built up, coders must be trained, and the responses must be coded into one of the categories before they can be tabulated and statistically analyzed. Compared to the simple process of tabulating the precoded responses to closed questions, the analysis of open-ended questions is complex and often troublesome.[15]

From the point of view of obtaining the information needed for a given investigation, however, each type of question has certain advantages and certain disadvantages. A closed question may help to ensure that the answers are given in a frame of reference that is relevant to the purpose of the inquiry and in a form that is usable in the analysis. For example, to the question, "About how often do you go to the movies, on the average?," if no alternative answers are supplied, one respondent might answer, "Not very often"; another, "When I have a date"; still another, "Only when there's something I especially want to see." If the investigator is interested in frequency of movie attendance, such answers are not usable. Provision of a check list with specific estimates ("more than once a week," "about once a week," "about three times a month," etc.) requires the respondent to frame his reply in terms that will be usable.

Sometimes the provision of alternative replies helps to make clear the meaning of the question. Respondents are more likely to understand the question, "Are you married, single, widowed, or divorced?" than the question, "What is your marital status?" This function of clarification may be important not only in relation to *words* whose meaning may not be generally known but in relation to *concepts* that may not be familiar to the respondent. A study by Gross, Mason, and McEachern (1958) of the school superintendency role provides an example. One aspect of this study concerned superintendents' perception of role conflicts (defined by the investigators as exposure to incompatible expectations on the part of different groups). Open-ended questions failed to bring replies relevant to the investigators' concept of "role conflict"; experimentation with different wording brought no success. Finally, they changed to a procedure of opening the interview with descriptions of situations involving problems that all superintendents face (criteria for the hiring and promotion of teachers, for example), including a number of fixed alternative actions that might be taken. The respondent was asked which of these actions various specified people or groups (e.g., his school board) would expect him to take. This procedure seemed to make the concept clear; the interviewers could then get meaningful answers to open-ended questions about *other* role-conflict situations that the superintendent had encountered.

---

[15] Footnote omitted.

A similar function of alternative responses is to make clear the *dimension* along which answers are sought. Consider the question, "Are you satisfied with your present wages?" One subject may answer, "No; I'd like to earn $100,000 a year." Another may say, "Yes, I think the wage scale at our plant is fair; I'm earning as much as other fellows who do the same kind of work in other places." The question involves neither words nor concepts that are difficult. But one subject has answered in terms of his level of aspiration (or fantasy), while the other has answered in terms of a judgment about the equity of his pay. To code the first respondent as dissatisfied and the second as satisfied might be misleading; if both of them had replied along the two dimensions, both might have said "No" in terms of level of aspiration but "Yes" in terms of fairness of wage scale. More precise wording of open-ended question might eliminate this difficulty by indicating more clearly which dimension was intended or by asking separately about both. However, the dimension along which answers are sought can frequently be indicated more clearly by a series of alternative responses than by the wording of the question itself.

Finally, the closed question may require the respondent himself to make a judgment about his attitude, rather than leaving this up to the interviewer or coder. This may or may not be desirable, depending on the nature of the question. On some issues the respondent may be in a better position to make the judgment. Suppose, in answer to the question, "How well satisfied are you with your job?," a respondent says: "Well, some things about it I like, some I don't. My boss is a nice guy; he doesn't chew us out for every little thing. And they're fair about promotions and things like that. But it's awfully dirty work, and it's a crummy old building; no decent place to eat, either." Let us say the plan of analysis calls for categorizing attitudes in terms of the following scale: definitely dissatisfied, more dissatisfied than satisfied, about half and half, more satisfied than dissatisfied, definitely satisfied. The coder may find it difficult to decide in which of the three middle categories to place this man. The man himself, however, might have little difficulty in making the judgment if he were presented with the alternative positions.

Most of these advantages of fixed-alternative questions have, however, corresponding disadvantages. One of the major drawbacks of the closed question is that it may force a statement of opinion on an issue about which the respondent does not have any opinion. Many individuals have no clearly formulated or crystallized opinions about many issues; this important characteristic is not likely to be revealed by a closed question. Inclusion of a "Don't know" alternative may help to provide an indication of a lack of crystallized opinion, but the tendency in much interviewing with questions of this sort is to press for a definite response and to accept a "Don't know" only as a last resort. Under such pressure, the answer chosen by a respondent may be an artifact of the specific wording or phrasing of the question or of the stated alternative responses. Suppose one were to ask, "Do you approve or disapprove of the Eisenhower Doctrine for aid to Middle Eastern countries threatened by Communist aggression?" It is easy to say "Approve" or "Disapprove," and many respondents may find this less embarrassing than admitting that they don't know what the Eisenhower Doctrine is, much less have an opinion about it. In the closed question, the reply is taken at face value. Open-ended questions, especially when they are used in an interview and can be followed by probes, provide a much better indication of whether the respondent has any information about the issue, whether he has a clearly formulated opinion about it, and how strongly he feels about it.

Even when a respondent has a clear opinion, a fixed-alternative question may not give an adequate representation of it because none of the choices corresponds exactly to his position, or because they do not allow for qualification. Take such a question as, "Which of the following considerations are most important to you in choosing a job? Interesting work; opportunity to assume responsibility; pleasant surroundings; congenial associates; opportunity for advancement; high salary; security. Place a *1* next to the one most important to you, 2 next to the one that is next most important, etc." Let us suppose that the items cover the range of relevant considerations for a given respondent, and that he has a fairly clear view. But his view may involve interrelations among the factors. In general, interesting work may be more important to him than a high salary. However, given a choice between two jobs, one of which pays twice as much as the other but is slightly less interesting, he may choose the higher-paying. Or there may be some lower limit of salary beyond which he feels he cannot afford to go, no matter how interesting the work. Such qualifications can be expressed in reply to an open-ended question; a closed question not only makes no provision for them, but even discourages the respondent from thinking about them.

Omission of possible alternative responses may lead to bias. Even when a space is provided for "other" replies, most respondents limit their answers to the alternatives provided. Omission of an alternative may seriously change the replies to even a factual question such as what magazines people read. In a study of applicants who were accepted by a certain college but did not actually enter, the subjects were presented with a check list of reasons for not attending. These reasons included such factors as the location of the college, its cost, the fact that it was not coeducational, the fact that specific desired courses were not offered, etc. However, the possibility that the applicant had in the end entered another college because it had a generally higher academic reputation was not included. Although a few respondents added this in the space provided for "other reasons," there was no way of estimating how many would have checked it had it been included among the suggested alternatives. Unless one can be reasonably certain, on the basis of either the logical possibilities or prior investigation, that the alternatives presented adequately cover the complete range of probable responses, it is safer to use an open-ended question, which does not bias the responses by suggesting some but not others.

The fact that the wording of questions is the same for all respondents may conceal the fact that different respondents make different interpretations, some of which may be quite different from those intended by the interviewer. This possibility exists, of course, in both closed and open questions, but it is much more likely to go undetected in the former. An instance of interpretations made from such varying frames of reference as to make the meaning of the obtained replies unclear has been reported by Crutchfield and Gordon (1947). A national survey used the following question: "After the war, would you like to see many changes or reforms made in the United States, or would you rather have the country remain pretty much the way it was before the war?" Most of the respondents replied that they wanted the country to remain "pretty much as it was." In a follow-up study, the same question was asked, but then followed by probes to ascertain what the respondents had in mind when they answered the question. The investigators identified seven different frames of reference: domestic issues (employment conditions, standards of living, etc.); technical improvements (better transportation, communications, etc.); political affairs; and so on. It seemed clear that no single interpretation of the responses to the closed question was justified.

From this discussion of the relative advantages and disadvantages of open and closed questions, it is apparent that the two differ in the purposes for which they are appropriate. Closed questions are more efficient where the possible alternative replies are known, limited in number, and clear-cut. Thus they are appropriate for securing factual information (age, education, home ownership, amount of rent, etc.) and for eliciting expressions of opinion about issues on which people hold clear opinions. Open-ended questions are called for when the issue is complex, when the relevant dimensions are not known, or when the interest of the research lies in the exploration of a *process* or of the individual's formulation of an issue. The closed question has the advantage of focusing the respondent's attention on the dimension of the problem in which the investigator is interested; by the same token, it does not provide information about the *respondent's own* formulation of the issue, the frame of reference in which he perceives it, the factors that are salient for him, the motivations that underlie his opinions. When these matters are the focus of interest, open-ended questions are essential.

Lazarsfeld (1944) has proposed that the development of a closed-question interview schedule be preceded by more intensive, freer interviews with a subsample of the population in order to discover the range of probable responses, the dimensions that are seen as relevant, and the various interpretations that may be made of the question wording. On the basis of such preliminary exploration, more meaningful closed questions can be formulated. He has also suggested another method of using the two types of question to supplement each other: after a survey using closed questions, more intensive interviews might be held with a subsample in order to delve more deeply into areas that appear significant. For many purposes, a combination of open and closed questions is most efficient; an interview or questionnaire need not consist entirely of one type or the other.

LESS STRUCTURED INTERVIEWS

For some research problems, a still more flexible approach than that provided by a standardized interview with open-ended questions is appropriate. Largely as a result of the influence of clinical interviewing and anthropological field work, a varied assortment of interviews has been developed in which neither the exact questions the interviewer asks nor the responses the subject is permitted to make are predetermined. Such interviews take various forms and go under various names—the "focused" interview, the "clinical" interview, the "depth" interview, the "nondirective" interview, etc. They are commonly used for a more intensive study of perceptions, attitudes, motivations, etc., than a standardized interview, whether with closed or open questions, permits. This type of interview is inherently more flexible, and of course it requires more skill on the part of the interviewer than do the standardized types. Obviously, this approach is impossible in a questionnaire.

The flexibility of the unstructured or partially structured interview, if properly used, helps to bring out the affective and value-laden aspects of the subject's responses and to determine the personal significance of his attitudes. Not only does it permit the subject's definition of the interviewing situation to receive full and detailed expression; it should also elicit the personal and social context of beliefs and feelings. This type of interview achieves its purpose to the extent that the subject's responses are spontaneous rather than forced, are highly specific and concrete rather than diffuse and general, are self-revealing and personal rather than superficial.

The freedom which the interviewer is permitted is, at once, both the major advantage and major disadvantage of interviews of this type.

The flexibility frequently results in a lack of comparability of one interview with another. Moreover, their analysis is more difficult and time-consuming than that of standardized interviews. There can be little doubt of their usefulness, in the hands of a skilled investigator, as a source of hypotheses that can later be submitted to a systematic test. Partially structured interviews are also used, on occasion, in studies *testing* hypotheses. However, the lack of comparability from interview to interview and the complexity of analysis usually make them less efficient for this purpose than standardized interviews.

Let us briefly discuss several of the major types of partially structured and unstructured interview.

In the *focused interview*, as described by Merton, Fiske, and Kendall (1956), the main function of the interviewer is to focus attention upon a given experience and its effects. He knows in advance what topics, or what aspects of a question, he wishes to cover. This list of topics or aspects is derived from his formulation of the research problem, from his analysis of the situation or experience in which the respondent has participated, and from hypotheses based on psychological or sociological theory. This list constitutes a framework of topics to be covered, but the manner in which questions are asked and their timing are left largely to the interviewer's discretion. He has freedom to explore reasons and motives, to probe further in directions that were unanticipated. Although the respondent is free to express completely his own line of thought, the direction of the interview is clearly in the hands of the interviewer. He wants definite types of information, and part of his task is to confine the respondent to discussion of the issues about which he wants knowledge.

Merton, Fiske, and Kendall (1956) have described this type of interview:

> First of all, the persons interviewed are known to have been involved in a *particular situation:* they have seen a film, heard a radio program, read a pamphlet, article or book, taken part in a psychological experiment or in an uncontrolled, but observed, social situation (for example, a political rally, a ritual or a riot). Secondly, the hypothetically significant elements, patterns, processes and total structure of this situation have been provisionally analyzed by the social scientist. Through this *content* or *sit-*

uational analysis, he has arrived at a set of hypotheses concerning the consequences of determinate aspects of the situation for those involved in it. On the basis of this analysis, he takes the third step of developing an *interview guide*, setting forth the major areas of inquiry and the hypotheses which provide criteria of relevance for the data to be obtained in the interview. Fourth and finally, the interview is focused on the subjective experiences of persons exposed to the pre-analyzed situation in an effort to ascertain *their definitions of the situation*. The array of reported responses to the situation helps test hypotheses and, to the extent that it includes unanticipated responses, gives rise to fresh hypotheses for more systematic and rigorous investigation.

The focused interview has been used effectively in the development of hypotheses about which aspects of a specific experience (a radio broadcast, a moving picture, a lecture, etc.) lead to changes in attitude on the part of those exposed to it. The interviewer, being equipped in advance with a content analysis of the stimulus experience, can usually distinguish the objective facts of the case from the subjective definitions of the situation. Thus, he is alerted to the possibility of "selective perception" and prepared to explore its implications. Suppose, for example, that one is concerned with reactions to a series of newspaper pictures portraying housing conditions in a slum neighborhood, intended for use in connection with a campaign for more stringent housing laws or for slum clearance and urban redevelopment. The pictures show broken stairs, wallpaper peeling off, holes in walls through which rats are reported to enter—in general, conditions that may reasonably be attributed to inadequate maintenance on the part of the landlord rather than slovenliness on the part of tenants. A respondent, in discussing the pictures, may say, "They show how these low-class people don't take care of their places; there's no use trying to give them decent housing, they just knock it to pieces anyway; you know, like they always say, if you give them bathtubs they just put coal in them." The interviewer, knowing that the content of the pictures is not intended to give this impression, can follow up the respondent's interpretation, trying to discover whether there are unconsidered aspects of the pictures that form a basis for this impression, whether it stems from the subject's stereotyped views, etc.

The definition of a focused interview may be broadened to include any interview in which the interviewer knows in advance what specific aspects of an experience he wishes to have the respondent cover in his discussion, whether or not the investigator has observed and analyzed the specific situation in which the respondent participated. For example, in a study of the functioning of a program of part-time work for high school students, one may prepare a set of questions to be covered even though he is not familiar with the specific job setting of each of the students. Such a list might include questions such as the following: "Does the student feel that he was given an adequate picture of the job before he started? Does he feel that his job is at a level appropriate to his skills?," etc.

Obviously, the more detailed the investigator's knowledge of the situation in which the person being interviewed has participated, and the more specific his hypotheses, the more precisely he can outline in advance the questions to be covered in the interview.

Somewhat similar to the focused interview is the *clinical interview*, the primary difference being that the clinical interview is concerned with broad underlying feelings or motivations, or with the course of the individual's life experiences, rather than with the effects of a specific experience. In this type of interview, too, the interviewer knows what aspects of feeling or experience he wants the respondent to talk about, but again the method of eliciting the information is left to his discretion. The "personal history" interview, used in social case work, prison administration, psychiatric clinics, and in social research using individual life histories, is perhaps the most common type of clinical interview. The specific aspects of the individual's life history which the interview is to cover are determined, as in all data-collection instruments, by the purpose for which the information is gathered.

For example, Lee (1957) was interested in the possibility that adolescents who become heroin addicts may be predisposed to addiction by family experiences that lead to certain personal characteristics. On the basis of earlier work with juvenile addicts, he and colleagues conducting related studies hypothesized that among boys living in the same neighborhood and thus exposed to roughly the same opportunities for using heroin, addicts are likely to differ from non-addicts in the following ways: they have relatively weak ego functioning, defective superego functioning, inadequate masculine identification, lack of realistic middle-class orientation, and distrust of major social institutions. Next, the investigators asked themselves what types of family environment might be expected to stimulate or enhance such characteristics. On the basis of theoretical considerations, largely drawn from psychoanalytic thinking, they constructed a list of circumstances or events of family life that might be expected to contribute to each of the five characteristics. For example, it was considered that factors such as the following might be conducive to weak ego functioning: inappropriate handling of childhood illnesses, discordant relationship between parents, the mother figure either passionate or hostile toward the boy, either parent's having unrealistically high or low aspirations for the boy, etc.

It seemed clear that relatively unstructured interviews would be a more appropriate method of getting the needed information than would a standardized series of questions. Accordingly, the interviewers visited the parents of the boys included in the study—a sample of addicts and a control group of non-addicts—and encouraged them to talk freely about their sons. The interviewers had no set questions to ask. They were instructed to cover the following major topics: the physical characteristics of the neighborhood and the house, the composition of the family and the household, the health history of the family, the present and early adolescent life situation of the subject, childhood training and socialization, relationships within the family, and relationships between the family and the "outside world." The interview guide indicated a number of subtopics to be covered under each of these major ones; for example, under "childhood training and socialization," the interviewer was to get information about early development, discipline and patterns of handling by parents, early socialization experiences, early school experiences, etc. Under each of these subtopics, the interview guide listed more specific points to be covered.

In the *nondirective interview*, the initiative is even more completely in the hands of the respondent. The term *nondirective* received its currency from a type of psychotherapy in which the patient is encouraged to express his feelings without directive suggestions or questions from the therapist. In a more limited sense, nondirection is implicit in most interviewing: that is, although the interviewer is expected to ask questions about a given topic, he is instructed not to bias or direct the respondent to one rather than another response.[16] In nondirective interviewing, however, the interviewer's function is simply to encourage the respondent to talk about a given topic with a minimum of direct questioning or guidance. He encourages the respondent to talk fully and freely by being alert to the feelings expressed in the statements of the respondent and by warm, but noncommittal, recognition of the subject's feelings. Perhaps the most typical remarks made by the interviewer in a nondirective interview are: "You feel that . . ." or "Tell me more" or "Why?" or "Isn't that interesting?" or, simply, "Uh huh."

The nondirective interviewer's function is primarily to serve as a catalyst to a comprehensive expression of the subject's feelings and beliefs and of the frame of reference within which his feelings and beliefs take on personal significance. To achieve this result, the interviewer must create a completely permissive atmosphere, in which the subject is free to express himself without fear of disapproval, admonition, or dispute, and without advice from the interviewer.[17]

---

[16] In "stress interviews," this statement does not hold. Here, the interviewer tries to see how well the respondent can function under the stress of baiting, disparagement, expressed hostility, etc. The term may also be applied to interviews in which the interviewer tries to see how much pressure or stress is required to change the respondent's expression of views.

[17] For a more detailed discussion of nondirective interviewing see Roethlisberger and Dickson (1939) and Rogers (1945).

POOR ORIGINAL COPY - BEST
AVAILABLE AT TIME FILMED

BIBLIOGRAPHY

BLANKENSHIP, A.B., 1946. How to conduct consumer and opinion research. Harper.

CRUTCHFIELD, R., and D.H. GORDON, 1947. Variations in respondents' interpretations of an opinion-poll question. International J.Opinion and Attitude Research, 1, 22-31.

EDWARDS, A.L., 1957a. Techniques of attitude scale construction. Appleton-Century-Crofts.

GROSS, N., W.S. MASON, and A.W. McEACHERN, 1958. Explorations in role analysis: studies of the school superintendency role. Wiley

KATZ, D., 1951. Social psychology and group process. Annual Review of Psychology, ed. by C. P. Stone. Annual Reviews, Inc. (Stanford, Calif.).

KINSEY, A.C., et al., 1948. Sexual behavior in the human male. Saunders.

KRAMER, B.M., 1949. Dimensions of prejudice. J. Psychology, 27, 389-451.

KRECH, D., and R.S. CRUTCHFIELD, 1948. Theory and problems of social psychology. McGraw-Hill.

LaPIERE, R.T., 1934. Attitudes vs. actions. Social Forces, 14, 230-237.

LAZARSFELD, P.F., 1935. The art of asking why. National Marketing Review, 1, 26-38.

LAZARSFELD, P.F., 1944. The controversy over detailed interviews--an offer for negotiation. Public Opinion Quarterly, 8, 38-60.

LEE, R.S., 1957. The family of the addict: a comparison of the family experiences of male juvenile heroin addicts and controls. Unpublished doctoral dissertation, New York University, Graduate School of Arts and Science.

MERTON, R.K., M. FISKE, and P.L. KENDALL, 1956. The focused interview. Free Press.

PACE, C.R., 1939. A situations test to measure social-political-economic attitudes. J.Social Psychology, 10, 331-344.

PARTEN, M.B., 1950. Surveys, polls, and samples. Harper.

RAPKIN, C., E. GRIER, and G. GRIER, 1957. Group relations in Newark--1957. Urban Research (New York).

ROGERS, C.R., 1945. The non-directive method as a technique for social research. American J. Sociology, 50, 279-283.

Bibliography

ROETHLISBERGER, F.J., and W.J. DICKSON, 1939. Management and the worker. Harvard University.

SELLTIZ, C., and associates, 1950. The acceptability of answers to anti-Semitic remarks. International J. Opinion and Attitude Research, 4, 353-390.

STOUFFER, S.A., 1949. An analysis of conflicting social norms. American Sociological Review, 14, 707-717.

ZEISEL, H., 1957. Say it with figures (4th ed.). Harper.

MAILED QUESTIONNAIRES

We have included the following reading about mailed questionnaires since they are sometimes a very useful device. We would caution, however, against their indiscriminate use. The mailed questionnaire is designed to be effective under a limited set of circumstances. As Professors Goode and Hatt indicate, "The appropriateness of the mailed questionnaire will depend upon the requirements of the research problem with regard to (1) the type of information required, (2) the type of respondent reached, (3) the accessibility of respondents, and (4) the precision of hypothesis."

The mailed questionnaire is obviously not to be used in a mechanical fashion, and it has many difficulties in measuring subtleties of behavior. Before it can be used effectively, therefore, the scientist must engage in preliminary investigation to probe and define his problem to the point of being able to construct hypotheses which can in fact be tested by a mailed questionnaire.

There is, however, more to the Goode and Hatt reading than a discussion of when to use a mailed questionnaire. The mailed questionnaire is self-administered and its effectiveness consequently is dependent upon its return. As the authors indicate, "Techniques for returning must be thought of as overlapping the techniques for eliciting responses," and they remind us that such pedestrian things as enclosing a self-addressed envelope and writing an accompanying letter legitimizing and explaining the relevance of one's study are necessary to increase the likelihood of the return of a questionnaire.

From William J. Goode and Paul K. Hatt, Methods in Social Research (New York: McGraw-Hill, 1952) pp. 170-183. Reprinted by permission.

William J. Goode and Paul K. Hatt, Methods in Social Research (New York: McGraw-Hill, 1952), pp. 170-183.

The preceding chapter dealt with the problem of formulating questions regardless of whether they were to be incorporated in a questionnaire, a schedule, or an interview guide. The questionnaire, it will be recalled, is differentiated from the schedule and interview guide by the fact that it is self-administered. Not all questionnaires are mailed. They may be administered to groups of people who have gathered together for any purpose. In this case, not all the problems which face the mailed questionnaire are present. Since, however, research of the group-questionnaire type differs from interview studies in the same direction, though not as sharply, as does the mailed-questionnaire technique, it seems best to discuss the latter in greatest detail. The student will be able to see for himself at what points the problems of the two are different.

In spite of many abuses, the mailed self-administering questionnaire remains a useful technique in sociological research. So long as this method is employed in appropriate research designs, it can frequently be rewarding. The crucial point in its use is determining whether or not this method of gathering data is the best one possible in the specific situation. The decision to use one method of collecting data over another method is complex and must take many factors into consideration.

THE RELATION OF THE MAILED QUESTIONNAIRE TO THE RESEARCH PROBLEM

The appropriateness of the mailed questionnaire will depend upon the requirements of the research problem with regard to (1) the type of information required, (2) the type of respondent reached, (3) the accessibility of respondents, and (4) the precision of the hypothesis.

The type of information required. First, very extensive bodies of data cannot often be secured through the use of the questionnaire. It is usually unwise to expect returns from a questionnaire which requires much more than 10 to 25 minutes to complete, and thus its use is restricted to rather narrow areas of data. Further, an extensive, survey type of problem suggests the need for so wide a range of data that personal participation and/or depth interviewing is required. It would, for example, have been impossible to secure the kind of information required by Whyte in his *Street Corner Society* by the use of questionnaires.[1]

Second, the questionnaire is effective only when the respondent is able or willing to express his reactions clearly. A considerable controversy has existed among social researchers as to whether the answers from the anonymous questionnaire are franker, or given with greater openness, because there is less fear when there is no immediate listener. This frankness has been contrasted with the interview situation, since it is claimed that the interviewer may inhibit the answers somewhat.[2]

This, however, seems to ignore the major problem. There can be no question that a good questionnaire can elicit frank answers on almost any subject, even such personal matters as sex and income. It is equally true, on the other hand, for the interview. When good questionnaires are compared with good interviews, it is likely that on the score of simple lying or frankness there is little to choose. The interviewer has the advantage of seeing the evasion, while the person who answers the questionnaire may feel less constraint in telling the truth.

The difference between the two is not in the dimension of frankness, but in that of *depth*. On this matter, there can be little doubt that the good interview can probe far more deeply than the best questionnaire. The reason for this can be seen in both statistical and social-relational terms. With reference to the former, the experienced researcher knows that the most simple question has a great depth of motivation, desire, attitude, and concrete complexity to it. In one project, for example, information was desired about the dating practices of divorced mothers. A simple question for a questionnaire might be, "How often do you date? Almost every night, two to six times a week, once a week, twice monthly, almost never?" However, such a question would not reach the complexity of observable patterns. Some women might have answered, "Almost never," because they were living with their ex-husbands, or with their fiancés, or with boarders. Further, each of the complex patterns breaks down into still further questions to be asked. A simple question to which there are eight possible answers, each of which must be probed by three possible alternatives, demands a series of 24 questions. These not only take up a large section of the space in any questionnaire, but would be very difficult to arrange spatially so that the respondent could thread his or her way through the maze.

At each deeper level of probing, the possible number of answers multiplies enormously, and all these possibilities must be included in the mailed questionnaire.

---

[1] William F. Whyte, *Street Corner Society* (Chicago: University of Chicago Press, 1943).

[2] Albert Ellis, "Questionnaire versus Interview Methods in the Study of Human Love Relationships," *American Sociological Review*, Vol. XII (1947), pp. 541–553, and Vol. XIII (1948), pp. 61–65.

The more important item for consideration, however, is that the *active work of the interviewer* is required to stimulate the respondent to express these deeper levels of motivation and feeling. It is not that in the interview it is possible to obtain a better rapport with the respondent, although it is likely that the interviewer can get and hold the emotional commitment of the respondent somewhat better. Rapport is not the most important element in this comparison, but it is rather the fact that the respondent is not *able* to give the necessary answers without help.

A social scientist attempted to compare the frankness of answers from the questionnaire and interview by asking comparable questions concerning the love experiences of a group of college girls.[3] Although in general the two techniques gave comparable results, with perhaps a slight advantage in favor of the questionnaire, one or two questions illustrated the present point, that frankness is not the sole issue. One dealt with whether the girl had ever been in love with two boys at once, and the interviews showed a substantially higher number of girls who admitted such attachments. The questions were phrased similarly, but it was discovered that when the interviewer received a "no" answer to this question, the further probe question was asked: "Never?" A number of girls then remembered such an attachment. However, there was no lack of frankness in the first answer. Rather, the probe question forced them to think again, to search their memories in order to be certain. Since the romantic complex in this country is opposed to such double attachments, many persons tend to forget them. This is also true for a number of incidents in the life histories of any individual. Even with the best of willingness to tell the truth, the truth does not come easily to the tongue.

A further point here is the fact that the questionnaire fails when its questions place an emotional burden on the respondent. This is a different dimension from that of being *unable* to answer because of unawareness of the complexity of one's motivation or the details of one's life history. The latter problem is met by the mental alertness of the interviewer, who probes when the answers are vague, meaningless, irrelevant, or contradictory. However, the face-to-face interview offers the *emotional support* of a sympathetic listener, in addition to the purely intellectual help. This factor may be of even greater significance in eliciting the facts than the latter. And, in any event, neither is obtainable when the respondent faces a mailed questionnaire.

The type of respondent reached. Although questionnaires have been used for a wide range of social groups, experience has shown that not all groups respond equally well. At the lowest level of differentiation, it must be clear that it is impossible to obtain a representative sample of the whole population by using this technique. A minimum necessity is the ability to read and write. Further, the amount of reading attention and writing skill required for most questionnaires is much greater than is implied by a minimum definition of "literate." For many who are able to read and write, neither burden is assumed willingly or easily.

---

[3] *Ibid.*

The consequence of this aspect of the questionnaire is that it cannot be used for a representative sample of the entire population. This does not detract from its usefulness in specific research situations, when more literate groups are the focus of the study. It is, however, only by recognizing the limitations of any research tool that we can utilize it to the fullest.[4]

Emphasis has been placed on the fact that a substantial segment of the population *cannot* answer a questionnaire. However, the *willingness* of the addressee to answer the questionnaire is a still further problem. The researcher is not there to add his own pleas to those which are printed or typewritten in the letter of introduction.

Both the "cannot" and the "will not" group bias the sample in a *known direction,* but to an *unknown degree.* Since questionnaire studies with a fairly wide population base have reported the percentage answering as from 20 per cent to around 70 per cent, it is clear that this type of sampling bias could be fatal to the validity of the study.

The direction of this bias is toward those who are interested in the subject matter, those who are higher in socioeconomic status, and those who have had more education. Thus, a questionnaire circulated among college professors of English, querying them about the kind of work load they carry, is likely to be answered by a high proportion of the respondents.

In one study, on traveling by train, a substantial proportion of returns was received from the first wave of questionnaires.[5] However, it is now standard procedure, in using the questionnaire, to send a second wave of questionnaires, or a third. Depending on the technique used, these can be sent to everyone on the primary list or only to those who have not answered. In this study, the second wave showed very different characteristics from the first. Looking more closely for the basis of the differences, it was seen that the first wave was mainly composed of those who had actually traveled on the newer types of "super" trains. The group had, therefore, a keener interest in questions about innovations in train travel and were willing to take the trouble of answering the questionnaire. It was only with the second wave, which stimulated those whose interest was less strong, that the peripheral group was tapped, those who had not enjoyed the newer train facilities.

[4] E. Jackson Baur, "Response Bias in a Mail Survey," *Public Opinion Quarterly,* Vol. XI (1947), pp. 594–600.

[5] Cf. Raymond F. Sletto, "Pretesting of Questionnaires," *American Sociological Review,* Vol. V (1940), pp. 193–200; and E. A. Suchman and B. McCandless, "Who Answers Questionnaires?" *Journal of Applied Psychology,* Vol. XXIV (1940), pp. 758–769.

The respondent will ordinarily not answer a questionnaire dealing with a subject with which he is unfamiliar, such as air travel, impending legislation, or experience with a particular branch of the government or an occupation. This lack of knowledge has been classed under the general heading of "lack of interest," for convenience, but it is clear that cases may exist in which these do not coincide at all. In a particular organization, for example, a controversy may have continued in a futile bickering fashion for many months, leaving the members willing to wash their hands of the matter and unwilling to answer a polling questionnaire which seems to bring up the issues once more. Another type of refusal to answer occurs when the questionnaire deals with the use of luxury goods, for in such cases there is a strong bias in the resulting sample, in the direction of those who have used the article in question.

The central point of mentioning these results is to underline the fact that the questionnaire is not an effective research tool for any but a highly select group of respondents. It is not effective because a biased sample is obtained. As noted in the section on sampling, the mere existence of a "bias" is not the primary point, but the fact that the extent of the bias is not measurable. If the exact characteristics of those who do not answer could be known, it would be possible to weight the results accordingly or to restrict the conclusions to the select group who responded.

8

On the other hand, questionnaires can be fairly effective if such a select group is the object of study. A group of business executives, for example, will answer a Dun and Bradstreet questionnaire with a low rate of refusals. The income-tax blank is a questionnaire and (backed in this case by police power) is fairly successful in eliciting a high rate of response. The *Dictionary of American Scholars* sends a questionnaire to those who are suggested for inclusion, and it is safe to say that the response is very close to 100 per cent.

Accessibility of respondents. Granted that the questionnaire is effective only for a highly select group of respondents, it is a very useful tool for certain situations in which the respondents are geographically widely dispersed. This, of course, actually reduces to a problem of time and money. The savings to be obtained from the use of the questionnaire are not to be formulated by saying that one is always cheaper than the other. To begin with, research on these tools has not proceeded far enough to be able to offer acceptable time and cost breakdowns. Actually, a large part of published research is done by academic or other scholars who, like home craftsmen or weekend farmers, work for the excitement of it. "Costs" include the actual outlay of money, with little allowance for their own time. Especially during a research project, the researcher and his associates are likely to pay little attention to the total number of hours worked. Furthermore, the extra costs on a questionnaire and an interview research are for different things. The questionnaire costs less per questionnaire than the interview costs per interview, but this is true only if immediate costs are considered. A great amount of time may be spent in waiting for successive questionnaire waves to come in. Or, the questionnaire may have to be supplemented by interviews. There may be extra transportation costs for interviewers, and extra costs for added efforts needed to fill in blanks which were left incomplete by the questionnaire respondents. In short, no simple comparison of costs can be made which will hold good in all cases.

However, it is certain that when the group of respondents is widely scattered, the costs will be less if the questionnaire is used. For example, if the researcher wishes to poll the membership of the American Sociological Society, transportation costs for interviewing would be excessive, both in money and time. Similarly, if an attitude study is to be made among the participants at a conference, there is not enough time to make the necessary interviews. However, questionnaires could be distributed to all those attending meetings at a particular hour. A selected group could be interviewed by questionnaire relative to a current controversy, even from a fairly isolated campus as a research base. Further, this could be done by a single researcher, without the large funds otherwise required to hire an interviewing staff, or to pay a national staff, such as the National Opinion Research Center, to carry out the interviewing in various cities.

This is not a blanket statement that the questionnaire is cheaper, only that it can be cheaper under certain circumstances. It must be kept clearly in mind that costs should not be computed on the basis of the *number* of interviews or questionnaires secured, but rather on the basis of the *amount of usable information* secured. For some information, the questionnaire cannot be used at all. For other information, under particular circumstances, the questionnaire is certainly cheaper than the interview and may be as adequate.

The precision of the hypothesis. From the preceding discussion, it is clear that the questionnaire is most useful when a considerable amount of exploratory work has narrowed the questions to be answered. The respondents are not likely to take the trouble to work out careful discriminations of attitude, personal history, and value. The questionnaire itself must do this, for the respondents to choose from. Of course, a great amount of exploratory work must go into any research project, no matter what tools are used. However, the interview remains a considerably more flexible instrument, even at the final stages. What is emphasized here is not the amount of preliminary work, but the sharpness of the hypothesis. The more closely focused the hypothesis, the more effective the questionnaire. At such a stage, the interview is equally applicable. However, the interview is effective even with a vaguely exploratory hypothesis. Indeed, the *testing of the questionnaire must utilize the interview.* A few questions can be tested by using the questionnaire without interviews. It can be determined, for example, that some of them are not answered or that the answers seem inconsistent. A few respondents will take the trouble to make marginal comments. However, this will not be adequate criticism. The questionnaire will have to be used with a selected sample who are then interviewed further concerning each question. Researchers often do this without selecting members of the respondent group, but the comments of colleagues or students are not sufficient at this stage. The meaning of each question can be adequately tested only by flexible, probing questioning of those who are representative of the final respondent group. This type of test cannot usually be done by the use of the questionnaire alone. As a consequence, it is only when the hypothesis has been rather sharply focused that the questionnaire is most effectively used.

SENDING OUT THE QUESTIONNAIRE

The questionnaire must be thought of as a kind of interview which is surrounded by peculiar obstacles. Consequently, many of the comments to be found in this section are equally applicable to the interview research technique. For example, it is clear that the respondent can judge the study only by what he can see. In interviewing, it is the interviewer himself. In the case of the questionnaire, it is several pieces of paper. These cannot adjust themselves to the situation. They are the same for all respondents. The questionnaire maker, then, must offer as impressive a presentation as possible, if the response is to be adequate. Only the papers are there to make his plea, and he cannot count on any personal charm or social skill when the respondent opens the envelope. He must, then, plan carefully and seek a great deal of professional help, before he sends out his queries. The following are concrete suggestions for building a questionnaire.

The appeal. A cover letter almost always accompanies the questionnaire. This is analogous to the opening "sales talk" of the interviewer, explaining what he is doing, why he is doing it, and for whom. The letter must explain these facts, but must cover as many objections as possible, since an unanswered objection means an ignored questionnaire. The introductory comments of the interviewer need not do so, since in most cases an elaborate explanation is unnecessary. The interviewer has answers ready for the objections, but he need not offer them unless the objections appear. The letter of appeal, however, must leave nothing unexplained.

On the other hand, it must be brief. Most of the recipients are not willing to read a long letter, and lengthiness usually destroys its impact.

Yet within it a number of basic facts must be presented. Some of these follow:

1. THE AUSPICES. Who is sanctioning the study? As in the case of the interview, this information must be given at the outset. If the research bureau is well known as a responsible, scientific group, the response is likely to be good. In some cases, the cover letter may devote a few sentences to explaining the character of the organization sanctioning the study, as well as the one carrying it out. The letterhead should convey the impression of scientific competence, and the address as well as the telephone number should be given to allow an easy check. Nothing should appear to be hidden or suspicious.

2. WHY THE STUDY? The interviewer may include this information in his introductory remarks, but it is often left out unless the respondent asks for an explanation. In the questionnaire cover letter, it is necessary to explain why the sanctioning organization needs the information, or why anyone at all wishes it. This need not be elaborate, but it must be sufficient to explain the need for answers to the questions. Questions dealing with marital relations can usually be explained adequately by referring to the modern "crisis of the family." Studies which explore religious behavior can allude to the secularism of the era. It cannot be assumed that *all* respondents will read the letter carefully, but it must be carefully tested to be certain that every phrase in it conveys the meaning intended. For *someone* will read the phrase, and an unfortunate interpretation will mean a loss of those respondents who found a different meaning in it.

3. WHY SHOULD THE RESPONDENT BOTHER TO ANSWER? Mention of the group authorizing the study, the group carrying it out, and the need for the information merely indicate that someone else is interested in the study. There must be an appeal to the respondent himself, which persuades him that *he* ought to participate. A great number of appeals have been used in the history of the questionnaire. Some have included money or promised it (usually no more than a nominal sum, such as 25 cents). Others have offered special services, such as marketing reports, radio bulletins, or samples of goods. Others have challenged the respondents to "prove what many thought impossible, that thousands of interested people would give information for this cause."

In general, the most effective appeal is an altruistic one. Although these special inducements such as money have increased the proportion of returns slightly, it remains true that no advantage that a research organization is likely to be able to offer will appeal to a large group of respondents. The amount of money that can be offered is trivial, and so are the other advantages, compared to the amount of time and thought requested. Whatever the student may believe concerning the cynicism of the age or the selfishness of people, extensive research has demonstrated that an appeal to disinterested motives is strongest. "The information is needed by thousands of leaders attempting to solve today's problems," "You will be contributing to the advancement of science," and "You will help improve the education of thousands of students who will attend Xiphosuran College in the future" are all better arguments than attempting to exploit the self-interest of the respondent.

The social scientist will ordinarily test which appeal is most effective for the group he is investigating, but the final form of his letter must *impel the respondent to action.* Mere presentation of the questionnaire is not enough. The impulse to activity is initially lacking, and the researcher is not even offering the respondent a sympathetic audience, as is done in the interview. The appeal is the "one-shot" stimulus, and its failure will result in the failure to gather data at all.

4. THE DIRECTIONS: HOW TO FILL IT OUT. Almost always, the beginning user of the questionnaire overrates the literacy of the average person. As anyone knows who has checked questionnaires filled out by civil service applicants, even those with a great deal of formal education are not always careful in following directions. Therefore, the student must make certain that a total stranger to the research can follow the directions with no effort. This aspect of the pretest should utilize individuals *with less education* than the group for which the directions are ultimately intended, to give a reasonable margin of error. It is not lack of education or intelligence which is the stumbling block, *but lack of attention and interest.* There is no primary good reason why motivation should be high. When high motivation cannot be expected, the demands on time and attention should be minimal. Directions, then, should be few and simple, both in the cover letter and in the questionnaire itself.

5. GUARANTEE OF ANONYMITY. One element which is used to back up the appeal is a guarantee of anonymity. Although many questionnaires do not ask for information which is embarrassing, respondents will ordinarily not answer if they have any reason to suspect that information about them will be made public. This suspicion will be more easily aroused, of course, if the information itself touches on delicate personal matters. The letter, then, should include a guarantee that the respondent will remain anonymous. There should be no request for names and no questions which are so detailed as to make identification easy (such as street addresses, etc.). The guarantee alone is not sufficient, if the questionnaire itself seems identifiable.

Such a rule is not to be interpreted too narrowly. When the investigating agency has a long history of making such studies within the same group, as an annual study of production from specified factories in an industry, such anonymity may become unnecessary.

William J. Goode and Paul K. Hatt

TECHNIQUES FOR FACILITATING THE RETURN OF THE QUESTIONNAIRE

As noted previously, the questionnaire is sometimes not mailed, but is administered to groups within the same room. This is the simplest collection situation. Usually, however, the document is mailed to the respondent, with directions to return it by mail. Perhaps the least effective method is having friends or students distribute copies to their own friends, with a request to return them after completion. Not only is the sample biased, but the proportion of loss is extremely high.

The questionnaire maker cannot seriously expect the respondent to exert himself in taking care of this task. Consequently, a stamped, self-addressed envelope should be sent with the questionnaire. There is some evidence that a regular stamp affixed to the envelope is more effective than postage from a mailing machine, or even than the stamped envelopes from the post office. The claim is that respondents do not like to see the stamps wasted. It is equally likely that the regular stamp gives some evidence that the organization is not carrying on a selling campaign. Commercial stamping is often associated with mailed advertisements, and there is a tendency on the part of many individuals to throw such envelopes away almost immediately.

As the student will see from these comments, "techniques for returning" must be thought of as overlapping the "techniques for eliciting response." Facilitating the return of the envelopes yields a greater response. Accurate typography, printed if possible, which is well spaced for easy reading, not only creates a favorable response in the recipient but also helps to eliminate errors in mailing. A slightly larger investment in format and typography will create a very great dividend in number of questionnaires completed and returned. Even the choice of paper is not to be taken lightly. The questionnaire will be handled by the recipient and the researcher, and it will be turned and shuffled many times while it is being coded and tabulated. The paper must, then, be of a good grade and of a convenient size. Many questionnaires, on the other hand, are reproduced on ordinary mimeographed paper, which is likely to smudge or blot when written upon with pencil or ink. As to typography, the researcher should investigate the different methods of reproduction before choosing. A number of new processes are much better than ordinary mimeographing, at only a slightly greater cost. Further, new processes are being developed as time goes on. One simple example of such a process, for the researcher with limited funds, is offset mimeographing, using bold Bodoni type such as is found on modern electric typewriters.

The result is very much like printing and considerably more impressive (and therefore more effective in eliciting responses) than ordinary typewriter type.

CHECKING THE SAMPLE

A major problem in using the questionnaire is the structure of the resulting sample. A low response is almost always indicative of a biased sample. However, a high proportion of questionnaires is not proof that no bias exists in the sample. If the questionnaire elicits a response from only certain groups among the respondents, the sample will be biased. To make certain that the resulting sample is representative, several devices have been used.

One of the most obvious is that of tabulating separately the respondents which result from (1) successive time periods, or (2) successive waves of questionnaires. For most studies, those who answer promptly are different from those who delay their answers. As noted previously, those who answer immediately have a much closer relationship to the subject matter, or wish to offer their opinions about a luxury product, or have a higher level of education. The student, then, may separate the answers received promptly from those received later. When there is very little difference between these two groups, and the percentage of response is high, he will have a fair assurance that the sampling bias is not great.

When successive waves of questionnaires are sent out, in order to obtain a very high level of response, such as 80 to 95 per cent, the same device can be used effectively. By an active follow-up campaign, a good response can be obtained from the group under study. Each new wave will refer to the previous mailing and comment on the importance of the work, appealing once more for an answer.

In order to check the resulting sample more carefully, some researchers have found it useful to assume that there is a final die-hard group who simply care little about the study and cannot be persuaded to bother greatly about it. For the final wave, then, a double postcard is sent, with a few basic questions to be marked with "yes," "no," or similarly brief answers. The double postcard is cheap, and this final group can be compared, for at least a few important characteristics (age, sex, occupation, etc.) with those who answered the questionnaire with little urging. This allows, then, some check on the sample obtained.

A comparable device, to be used with successive questionnaire waves, eliminates some of the cost of duplication. This technique utilizes a small section of the questionnaire, or a separate sheet, on which the respondent can put his name, either with or without cross-tabulation data such as age, sex, etc. This identifying tab is mailed separately by the respondent. The duplicate, stamped envelopes are an added cost, but on receipt of the tab the student can save the cost of sending a copy of the second wave to that respondent. The system preserves the anonymity of the questionnaire, while allowing the researcher to know who has answered the questions.

The further advantage is an obvious one; the unknown character of the sample bias is thereby clarified. The researcher usually knows something about his respondents, else he would not be sending his questionnaire to the particular persons concerned. Even a few items of information allow a simple comparison between the answerers and the nonanswerers. To the extent, then, that any such comparisons can be made, the extent of the bias can be known.

Knowing the direction of the bias, or something about it, may lead to a decision to *weight* the tabulations, under the assumption that those who do not answer have the same characteristics as those who answer very late. A more sophisticated technique was developed by the Bureau of the Census, which involves the use of face-to-face interviews for a certain percentage of those who do not answer the questionnaire.[6] Given the *cost* of such a field interview as compared to the cost of mailing the questionnaire, and the *degree of uncertainty* in the sample structure, which is increased as the percentage of response is lower, it is possible to calculate
8 how many interviews ought to be used for any given level of response. It is always assumed, of course, that the *level of precision is not to be changed*. If the answers are not valid and reliable, that is, adequate for the problem, there is little point in doing the study at all. Hansen and Hurwitz have actually calculated hypothetical tables from which the student can determine how large a proportion of interviews to obtain, given the factors of cost, percentage of response, etc.[7]

Although the beginning student is not likely to be able to use such a technique, it is of great utility in large surveys where a definite level of precision is demanded. The interviews allow a clear picture to be drawn of those who do not answer the questionnaire. In this way, not only is the *direction* of the sampling bias known, but the *degree* and *value* of that bias. Consequently, the resulting tabulations can be weighted with accuracy. On the other hand, not all the savings of the questionnaire technique are lost unless the level of response is very low.

[6] Morris H. Hansen and William N. Hurwitz, "The Problem of Non-response in Sample Surveys," *Journal of the American Statistical Association*, Vol. XLI (1946), pp. 517–529. This is an adaptation of Neyman's principle of double sampling—J. Neyman, "Contributions to the Theory of Sampling Human Populations," *Journal of the American Statistical Association*, Vol. XXV (1938), pp. 101–116—and was used at an early date in sampling lumber production.

[7] Hansen and Hurwitz, *op. cit.*, pp. 521, 523–524.

OPEN-ENDED QUESTIONS

Although the previous remarks have suggested that the most effective questions for the self-administered questionnaire are those which are highly *structured, i.e.,* posing all the possible alternatives for the respondent to choose from, the *unstructured* question may yield further information. Presumably, the work prior to the final formulation of the questionnaire has narrowed the possibilities somewhat, but questions allowing greater freedom to the respondent may still be required. It must be pointed out, however, that such questions are useful with an even smaller proportion of the total population, for such questions require a higher standard of literacy. All the previous remarks on this score are pertinent, with the further emphasis that the open-ended question demands a still greater amount of energy, willingness, and care on the part of the respondent.

As a consequence, unless the sample is extremely carefully chosen and carefully restricted to a rather literate group, the answers to such free questions are likely to be incomplete, couched in clichés, or nonexistent. The problem of coding them is, of course, no different for the questionnaire than for other data-gathering techniques.

If, however, such questions are to be used, the researcher must be rather generous with space on the sheet. A space which is only large enough for a sentence is likely to be filled with no more than a cryptic phrase. It is far better to allow the respondent more space than seems necessary, as a stimulus to a fairly full response.

SUMMARY

The mailed questionnaire has been very useful in social research in the past and is likely to continue to be so in the future. When the student chooses between the mailed questionnaire and the face-to-face interview, he must weight their relative advantages for the purposes and conditions of the project. The questionnaire does not allow complex probe questions which require the respondent to thread his way through many levels of subquestions. Further, the interviewer is not there to give emotional support and other stimuli to increase the respondent's ability and willingness to answer. The questionnaire can be most fruitfully used for highly select respondents with a strong interest in the subject matter, greater education, and higher socioeconomic status. Otherwise, an unknown amount of bias is introduced into the sample. However, when respondents are widely dispersed geographically or will be in one spot only briefly, the questionnaire may be very useful. Whether its cost, for the information needed, is less will depend upon many factors such as degree of response, costs of mailing successive waves, waiting time, etc. Finally, the questionnaire is efficient only when its basic hypothesis is relatively precise. It is not a useful device for gathering preliminary observations concerning social behavior. The present chapter has attempted to analyze the factors which should be weighed in choosing between the interview and the mailed questionnaire and has also described the procedures to be followed in using the questionnaire.

SUGGESTED READINGS

Edgerton, Harold A., Steuart H. Britt, and Ralph D. Norman, "Objective Differences among Various Types of Respondents to a Mailed Questionnaire," *American Sociological Review*, Vol. XII (1947), pp. 435–444.

Parten, Mildred, *Surveys, Polls, and Samples: Practical Procedures* (New York: Harper, 1950), Chap. 11.

Reuss, Carl F., "Differences between Persons Responding and Not Responding to a Mailed Questionnaire," *American Sociological Review*, Vol. VIII, pp. 433–438.

Sletto, Raymond F., "Pretesting of Questionnaires," *American Sociological Review*, Vol. V (April, 1940), pp. 193–200.

PARTICIPANT OBSERVATION

There is a common theme which runs through the two preceding readings and the following selection of Professor Becker's: all three readings refer to the function of participant observation in social science research. All three writers agree that the use of participant observation as a technique of data collection is most useful in what Becker refers to as acquiring an "understanding of a particular organization or substantive problem rather than demonstrating relationships between abstractly defined variables . . . ."

The significance of the Becker reading is found not only in its detailed discussion of participant observation but also its analysis of the distinction between qualitative and quantitative evidence in social science research. Indeed, it is the qualitative aspect of participant observation which is both its strength and its weakness. On the one hand, participant observation is the most human aspect of social science research, in a sense, in that it is heavily dependent upon the intuition and creativity of the researcher. On the other hand, however, its use can be criticized on the grounds that "the data of participant observation do not lend themselves to . . . ready summary." In other words, the data frequently consist of many different kinds of observations which cannot, as Becker points out, be "simply categorized and counted without losing some of their value as evidence."

What Becker does in the following reading, essentially, is to present an argument for the reliability of qualitative data. He tells us that participant observation involves a sequence of analytical analysis as well as a systematic scrutinization of the collected facts. There are, he says: "Three distinct stages of analysis conducted in the field itself." The participant observer selects and defines problems, concepts, and indices: checks the frequency and distribution of phenomena; and incorporates individual findings into a model of the phenomenon under study. In addition, the researcher goes through a process of evaluating reported testimony. For example, he inquires into the role of the respondent within the institution being studied and then proceeds to check his responses against others playing the same role. Further, the participant observer must also evaluate his own role in relation to the reliability of the data collected. What all of this means is that we must build a basis for judging the value of qualitative data--and the basis must be as obvious and systematic as possible.

From Howard S. Becker, "Problems of Inference and Proof in Participant Observation," American Sociological Review, 23 (December, 1958), pp. 652-660. Reprinted by permission.

Howard S. Becker, "Problems of Inference and Proof in Participant Observation," American Sociological Review, XXIII (December 1958), pp. 652-660.

THE participant observer gathers data by participating in the daily life of the group or organization he studies.[1] He watches the people he is studying to see what situations they ordinarily meet and how they behave in them. He enters into conversation with some or all of the participants in these situations and discovers their interpretations of the events he has observed.

Let me describe, as one specific instance of observational technique, what my colleagues and I have done in studying a medical school. We went to lectures with students taking their first two years of basic science and frequented the laboratories in which they spend most of their time, watching them and engaging in casual conversation as they dissected cadavers or examined pathology specimens. We followed these students to their fraternity houses and sat around while they discussed their school experiences. We accompanied students in the clinical years on rounds with attending physicians, watched them examine patients on the wards and in the clinics, sat in on discussion groups and oral exams. We ate with the students and took night call with them. We pursued释 internes and residents through their crowded schedules of teaching and medical work. We stayed with one small group of students on each service for periods ranging from a week to two months, spending many full days with them. The observational situations allowed time for conversation and we took advantage of this to interview students about things that had happened and were about to happen, and about their own backgrounds and aspirations.

Sociologists usually use this method when they are especially interested in understanding a particular organization or substantive problem rather than demonstrating relations

---

* This paper developed out of problems of analysis arising in a study of a state medical school. The study is sponsored by Community Studies, Inc., of Kansas City, Missouri. It is directed by Everett C. Hughes; Anselm Strauss is also a member of the research team. Most of the material presented here has been worked out with the help of Blanche Geer, who has been my partner in field work and analysis in this study. I am grateful to Alvin W. Gouldner for a thorough critique of an earlier draft.

Substantive papers on the study, whose findings are made use of throughout, include: Howard S. Becker and Blanche Geer, "The Fate of Idealism in Medical School," *American Sociological Review*, 23 (February, 1958), pp. 50–56, and "Student Culture in Medical School," *Harvard Educational Review*, 28 (Winter, 1958), pp. 70–80. Another paper on participant observation by the same authors is "Participant Observation and Interviewing: A Comparison," *Human Organization*, 16 (Fall, 1957), pp. 28–32.

[1] There is little agreement on the specific referent of the term *participant observation*. See Raymond L. Gold, "Roles in Sociological Field Observations," *Social Forces*, 36 (March, 1958), pp. 217–223, for a useful classification of the various procedures that go by this name. Our own research, from which we have drawn our illustrations, falls under Gold's type, "participant-as-observer." The basic methods discussed here, however, would appear to be similar in other kinds of field situations.

between abstractly defined variables. They attempt to make their research theoretically meaningful, but they assume that they do not know enough about the organization *a priori* to identify relevant problems and hypotheses and that they must discover these in the course of the research. Though participant observation can be used to test *a priori* hypotheses, and therefore need not be as unstructured as the example I have given above, this is typically not the case. My discussion refers to the kind of participant observation study which seeks to discover hypotheses as well as to test them.

Observational research produces an immense amount of detailed description; our files contain approximately five thousand single-spaced pages of such material. Faced with such a quantity of "rich" but varied data, the researcher faces the problem of how to analyze it systematically and then to present his conclusions so as to convince other scientists of their validity. Participant observation (indeed, qualitative analysis generally) has not done well with this problem, and the full weight of evidence for conclusions and the processes by which they were reached are usually not presented, so that the reader finds it difficult to make his own assessment of them and must rely on his faith in the researcher.

In what follows I try to pull out and describe *the basic analytic operations carried on in participant observation,* for three reasons: to make these operations clear to those unfamiliar with the method; by attempting a more explicit and systematic description, to aid those working with the method in organizing their own research; and, most importantly, in order to propose some changes in analytic procedures and particularly in reporting results which will make the processes by which conclusions are reached and substantiated more accessible to the reader.

The first thing we note about participant observation research is that analysis is carried on *sequentially,*[2] important parts of the analysis being made while the researcher is still gathering his data. This has two obvious consequences: further data gathering takes its direction from provisional analyses; and the amount and kind of provisional analysis carried on is limited by the exigencies of the field work situation, so that final comprehensive analyses may not be possible until the field work is completed.

We can distinguish three distinct stages of analysis conducted in the field itself, and a fourth stage, carried on after completion of the field work. These stages are differentiated, first, by their logical sequence: each succeeding stage depends on some analysis in the preceding stage. They are further differentiated by the fact that different kinds of conclusions are arrived at in each stage and that these conclusions are put to different uses in the continuing research. Finally, they are differentiated by the different criteria that are used to assess evidence and to reach conclusions in each stage. The three stages of field analysis are: the selection and definition of problems, concepts, and indices; the check on the frequency and distribution of phenomena; and the incorporation of individual findings into a model of the organization under study.[3] The fourth stage of final analysis involves problems of presentation of evidence and proof.

SELECTION AND DEFINITION OF PROBLEMS, CONCEPTS, AND INDICES

In this stage, the observer looks for problems and concepts that give promise of yielding the greatest understanding of the organization he is studying, and for items which may serve as useful indicators of facts which are harder to observe. The typical conclusion that his data yield is the simple one that a given phenomenon exists, that a certain event occurred once, or that two phenomena were observed to be related in one instance; the conclusion says nothing about the frequency or distribution of the observed phenomenon.

By placing such an observation in the context of a sociological theory, the observer selects concepts and defines problems for further investigation. He constructs a theoretical model to account for that one case, intending to refine it in the light of subsequent findings. For instance, he might find the following: "Medical student X referred to one of his patients as a 'crock' today."[4] He may then connect this finding with a sociological theory suggesting that occupants of one social category in an institution classify members of other categories by criteria derived from the kinds of problems these other persons raise in the relationship. This combination of observed fact and theory directs him to look for the problems in student-patient interaction indicated by the term "crock." By discovering specifically what students have in mind in using the term, through questioning and continued observation, he may develop specific hypotheses about the nature of these interactional problems.

---

[2] In this respect, the analytic methods I discuss bear a family resemblance to the technique of *analytic induction.* Cf. Alfred Lindesmith, *Opiate Addiction* (Bloomington: Principia Press, 1947), especially pp. 5–20, and the subsequent literature cited in Ralph H. Turner, "The Quest for Universals in Sociological Research," *American Sociological Review,* 18 (December, 1953), pp. 604–611.

[3] My discussion of these stages is abstract and simplified and does not attempt to deal with practical and technical problems of participant observation study. The reader should keep in mind that in practice the research will involve all these operations simultaneously with reference to different particular problems.

Conclusions about a single event also lead the observer to decide on specific items which might be used as indicators[5] of less easily observed phenomena. Noting that in at least one instance a given item is closely related to something less easily observable, the researcher discovers possible shortcuts easily enabling him to observe abstractly defined variables. For example, he may decide to investigate the hypothesis that medical freshmen feel they have more work to do than can possibly be managed in the time allowed them. One student, in discussing this problem, says he faces so much work that, in contrast to his undergraduate days, he is forced to study many hours over the weekend and finds that even this is insufficient. The observer decides, on the basis of this one instance, that he may be able to use complaints about weekend work as an indicator of student perspectives on the amount of work they have to do. The selection of indicators for more abstract variables occurs in two ways: the observer may become aware of some very specific phenomenon first and later see that it may be used as an indicator of some larger class of phenomena; or he may have the larger problem in mind and search for specific indicators to use in studying it.

Whether he is defining problems or selecting concepts and indicators, the researcher at this stage is using his data only to speculate about possibilities. Further operations at later stages may force him to discard most of the provisional hypotheses. Nevertheless, problems of evidence arise even at this point, for the researcher must assess the individual items on which his speculations are based in order not to waste time tracking down false leads. We shall eventually need a systematic statement of canons to be applied to individual items of evidence. Lacking such a statement, let us consider some commonly used tests. (The observer typically applies these tests as seems reasonable to him during this and the succeeding stage in the field. In the final stage, they are used more systematically in an overall assessment of the total evidence for a given conclusion.)

The Credibility of Informants. Many items of evidence consist of statements by members of the group under study about some event which has occurred or is in process. Thus, medical students make statements about faculty behavior which form part of the basis for conclusions about faculty-student relations. These cannot be taken at face value; nor can they be dismissed as valueless. In the first place, the observer can use the statement as evidence *about the event*, if he takes care to evaluate it by the criteria an historian uses in examining a personal document.[6] Does the informant have reason to lie or conceal some of what he sees as the truth? Does vanity or expediency lead him to mis-state his own role in an event or his attitude toward it? Did he actually have an opportunity to witness the occurrence he describes or is hearsay the source of his knowledge? Do his feelings about the issues or persons under discussion lead him to alter his story in some way?

---

[4] The examples of which our hypothetical observer makes use are drawn from our own current work with medical students.

[5] The problem of indicators is discussed by Paul F. Lazarsfeld and Allen Barton, "Qualitative Measurement in the Social Sciences: Classification, Typologies, and Indices," in Daniel Lerner and Harold D. Lasswell, editors, *The Policy Sciences: Recent Developments in Scope and Method,* Stanford: Stanford University Press, 1951, pp. 155–192; "Some Functions of Qualitative Analysis in Sociological Research," *Sociologica,* 1 (1955), pp. 324–361 (this important paper parallels the present discussion in many places); and Patricia L. Kendall and Paul F. Lazarsfeld, "Problems of Survey Analysis," in R. K. Merton and P. F. Lazarsfeld, editors, *Continuities in Social Research,* Glencoe: Free Press, 1950, pp. 183–186.

[6] Cf. Louis Gottschalk, Clyde Kluckhohn, and Robert Angell, *The Use of Personal Documents in History, Anthropology, and Sociology,* New York: Social Science Research Council, 1945, pp. 15–27, 38–47.

Secondly, even when a statement examined in this way proves to be seriously defective as an accurate report of an event, it may still provide useful evidence for a different kind of conclusion. Accepting the sociological proposition that an individual's statements and descriptions of events are made from a perspective which is a function of his position in the group, the observer can interpret such statements and descriptions as indications of the individual's perspective on the point involved.

Volunteered or Directed Statements. Many items of evidence consist of informants' remarks to the observer about themselves or others or about something which has happened to them; these statements range from those which are a part of the running casual conversation of the group to those arising in a long intimate tete-a-tete between observer and informant. The researcher assesses the evidential value of such statements quite differently, depending on whether they have been made independently of the observer (volunteered) or have been directed by a question from the observer. A freshman medical student might remark to the observer or to another student that he has more material to study than he has time to master; or the observer might ask, "Do you think you are being given more work than you can handle?", and receive an affirmative answer.

This raises an important question: to what degree is the informant's statement the same one he might give, either spontaneously or in answer to a question, in the absence of the observer? The volunteered statement seems likely to reflect the observer's preoccupations and possible biases less than one which is made in response to some action of the observer, for the observer's very question may direct the informant into giving an answer which might never occur to him otherwise. Thus, in the example above, we are more sure that the students are concerned about the amount of work given them when they mention this of their own accord than we are when the idea may have been stimulated by the observer asking the question.

The Observer-Informant-Group Equation. Let us take two extremes to set the problem. A person may say or do something when alone with the observer or when other members of the group are also present. The evidential value of an observation of this behavior depends on the observer's judgment as to whether the behavior is equally likely to occur in both situations. On the one hand, an informant may say and do things when alone with the observer that accurately reflect his perspective but which would be inhibited by the presence of the group. On the other hand, the presence of others may call forth behavior which reveals more accurately the person's perspective but would not be enacted in the presence of the observer alone. Thus, students in their clinical years may express deeply "idealistic" sentiments about medicine when alone with the observer, but behave and talk in a very "cynical" way when surrounded by fellow students. An alternative to judging one or the other of these situations as more reliable is to view each datum as valuable in itself, but with respect to different conclusions. In the example above, we might conclude that students have "idealistic" sentiments but that group norms may not sanction their expression.[7]

In assessing the value of items of evidence, we must also take into account the observer's role in the group. For the way the subjects of his study define that role affects what they will tell him or let him see. If the observer carries on his research incognito, participating as a full-fledged member of the group, he will be privy to knowledge that would normally be shared by such a member and might be hidden from an outsider. He could properly interpret his own experience as that of a hypothetical "typical" group member. On the other hand, if he is known to be a researcher, he must learn how group members define him and in particular whether or not they believe that certain kinds of information and events should be kept hidden from him. He can interpret evidence more accurately when the answers to these questions are known.

---

[7] See further, Howard S. Becker, "Interviewing Medical Students," *American Journal of Sociology*, 62 (September, 1956), pp. 199–201.

CHECKING THE FREQUENCY AND DISTRIBUTION OF PHENOMENA

The observer, possessing many provisional problems, concepts, and indicators, now wishes to know which of these are worth pursuing as major foci of his study. He does this, in part, by discovering if the events that prompted their development are typical and widespread, and by seeing how these events are distributed among categories of people and organizational sub-units. He reaches conclusions that are essentially quantitative, using them to describe the organization he is studying.

Participant observations have occasionally been gathered in standardized form capable of being transformed into legitimate statistical data.[8] But the exigencies of the field usually prevent the collection of data in such a form as to meet the assumptions of statistical tests, so that the observer deals in what have been called "quasi-statistics."[9] His conclusions, while implicitly numerical, do not require precise quantification. For instance, he may conclude that members of freshmen medical fraternities typically sit together during lectures while other students sit in less stable smaller groupings. His observations may indicate such a wide disparity between the two groups in this respect that the inference is warranted without a standardized counting operation. Occasionally, the field situation may permit him to make similar observations or ask similar questions of many people, systematically searching for quasi-statistical support for a conclusion about frequency or distribution.

In assessing the evidence for such a conclusion the observer takes a cue from his statistical colleagues. Instead of arguing that a conclusion is either totally true or false, he decides, if possible, how *likely* it is that his conclusion about the frequency or distribution of some phenomenon is an accurate quasi-statistic, just as the statistician decides, on the basis of the varying values of a correlation coefficient or a significance figure, that his conclusion is more or less likely to be accurate. The kind of evidence may vary considerably and the degree of the observer's confidence in the conclusion will vary accordingly. In arriving at this assessment, he makes use of some of the criteria described above, as well as those adopted from quantitative techniques.

Suppose, for example, that the observer concludes that medical students share the perspective that their school should provide them with the clinical experience and the practice in techniques necessary for a general practitioner. His confidence in the conclusion would vary according to the nature of the evidence, which might take any of the following forms: (1) *Every* member of the group said, *in response to a direct question,* that this was the way he looked at the matter. (2) *Every* member of the group *volunteered* to an observer that this was how he viewed the matter. (3) *Some given proportion* of the group's members either *answered* a direct question or *volunteered* the information that he shared this perspective, but none of the others was asked or volunteered information on the subject. (4) Every member of the group was asked or volunteered information, but *some given proportion said* they viewed the matter from the differing perspective of a prospective specialist. (5) No one was asked questions or volunteered information on the subject, but *all members were observed to engage in behavior* or to make other statements from which the analyst *inferred* that the general practitioner perspective was being used by them as a basic, though unstated, premise. For example, all students might have been observed to complain that the University Hospital received too many cases of rare diseases that general practitioners rarely see. (6) *Some given proportion* of the group *was observed* using the general practitioner perspective as a basic premise in their activities, but *the rest of the group* was not observed engaging in such activities. (7) *Some proportion* of the group *was observed* engaged in activities implying the general practitioner perspective while *the remainder* of the group was observed engaged in activities implying the perspective of the prospective specialist.

---

[8] See Peter M. Blau, "Co-operation and Competition in a Bureaucracy," *American Journal of Sociology,* 59 (May, 1954), pp. 530–535.

[9] See the discussion of quasi-statistics in Lazarsfeld and Barton, "Some Functions of Qualitative Analysis . . .," *op. cit.,* pp. 346–348.

The researcher also takes account of the possibility that his observations may give him evidence of different kinds on the point under consideration. Just as he is more convinced if he has many items of evidence than if he has a few, so he is more convinced of a conclusion's validity if he has *many kinds* of evidence.[10] For instance, he may be especially persuaded that a particular norm exists and affects group behavior if the norm is not only described by group members but also if he observes events in which the norm can be "seen" to operate—if, for example, students tell him that they are thinking of becoming general practitioners and he also observes their complaints about the lack of cases of common diseases in University Hospital.

The conclusiveness which comes from the convergence of several kinds of evidence reflects the fact that separate varieties of evidence can be reconceptualized as deductions from a basic proposition which have now been verified in the field. In the above case, the observer might have deduced the desire to have experience with cases like those the general practitioner treats from the desire to practice that style of medicine. Even though the deduction is made after the fact, confirmation of it buttresses the argument that the general practitioner perspective is a group norm.

It should be remembered that these operations, when carried out in the field, may be so interrupted because of imperatives of the field situation that they are not carried on as systematically as they might be. Where this is the case, the overall assessment can be postponed until the final stage of postfield work analysis.

CONSTRUCTION OF SOCIAL SYSTEM MODELS

The final stage of analysis in the field consists of incorporating individual findings into a generalized model of the social system or organization under study or some part of that organization.[11] The concept of social system is a basic intellectual tool of modern sociology. The kind of participant observation discussed here is related directly to this concept, explaining particular social facts by explicit reference to their involvement in a complex of interconnected variables that the observer constructs as a theoretical model of the organization. In this final stage, the observer designs a descriptive model which best explains the data he has assembled.

The typical conclusion of this stage of the research is a statement about a set of complicated interrelations among many variables. Although some progress is being made in formalizing this operation through use of factor analysis and the relational analysis of survey data,[12] observers usually view currently available statistical techniques as inadequate to express their conceptions and find it necessary to use words. The most common kinds of conclusions at this level include:

(1) Complex statements of the necessary and sufficient conditions for the existence of some phenomenon. The observer may conclude, for example, that medical students develop consensus about limiting the amount of work they will do because (a) they are faced with a large amount of work, (b) they engage in activities which create communication channels between all members of the class, and (c) they face immediate dangers in the form of examinations set by the faculty.

(2) Statements that some phenomenon is an "important" or "basic" element in the organization. Such conclusions, when elaborated, usually point to the fact that this phenomenon exercises a persistent and continuing influence on diverse events. The observer might conclude that the ambition to become a

---

[10] See Alvin W. Gouldner, *Patterns of Industrial Bureaucracy*, Glencoe, Ill.: Free Press, 1954, pp. 247–269.

[11] The relation between theories based on the concept of social system and participant observation was pointed out to me by Alvin W. Gouldner. See his "Some Observations on Systematic Theory, 1945–55," in Hans L. Zetterberg, editor, *Sociology in the United States of America*, Paris: UNESCO, 1956, pp. 34–42; and "Theoretical Requirements of the Applied Social Sciences," *American Sociological Review*, 22 (February, 1957), pp. 92–102.

[12] See Alvin W. Gouldner, "Cosmopolitans and Locals: Toward an Analysis of Latent Social Roles," *Administrative Science Quarterly*, 2 (December, 1957), pp. 281–306, and 3 (March, 1958), pp. 444–480; and James Coleman, "Relational Analysis: The Study of Social Structure with Survey Methods," mimeographed.

general practitioner is "important" in the medical school under study, meaning that many particular judgments and choices are made by students in terms of this ambition and many features of the school's organization are arranged to take account of it.

(3) Statements identifying a situation as an instance of some process or phenomenon described more abstractly in sociological theory. Theories posit relations between many abstractly defined phenomena, and conclusions of this kind imply that relationships posited in generalized form hold in this particular instance. The observer, for example, may state that a cultural norm of the medical students is to express a desire to become a general practitioner; in so doing, he in effect asserts that the sociological theory about the functions of norms and the processes by which they are maintained which he holds to be true in general is true in this case.

In reaching such types of conclusions, the observer characteristically begins by constructing models of parts of the organization as he comes in contact with them, discovers concepts and problems, and the frequency and distribution of the phenomena these call to his attention. After constructing a model specifying the relationships among various elements of this part of the organization, the observer seeks greater accuracy by successively refining the model to take account of evidence which does not fit his previous formulation; [13] by searching for negative cases (items of evidence which run counter to the relationships hypothesized in the model) which might force such revision; and by searching intensively for the interconnections *in vivo* of the various elements he has conceptualized from his data. While a provisional model may be shown to be defective by a negative instance which crops up unexpectedly in the course of the field work, the observer may infer what kinds of evidence would be likely to support or to refute his model and may make an intensive search for such evidence.[14]

After the observer has accumulated several partial-models of this kind, he seeks connections between them and thus begins to construct an overall model of the entire organization. An example from our study shows how this operation is carried on during the period of field work. (The reader will note, in this example, how use is made of findings typical of earlier stages of analysis.)

When we first heard medical students apply the term "crock" to patients we made an effort to learn precisely what they meant by it. We found, through interviewing students about cases both they and the observer had seen, that the term referred in a derogatory way to patients with many subjective symptoms but no discernible physical pathology. Subsequent observations indicated that this usage was a regular feature of student behavior and thus that we should attempt to incorporate this fact into our model of student-patient behavior. The derogatory character of the term suggested in particular that we investigate the reasons students disliked these patients. We found that this dislike was related to what we discovered to be the students' perspective on medical school: the view that they were in school to get experience in recognizing and treating those common diseases most likely to be encountered in general practice. "Crocks," presumably having no disease, could furnish no such experience. We were thus led to specify connections between the student-patient relationship and the student's view of the purpose of his professional education. Questions concerning the genesis of this perspective led to discoveries about the organization of the student body and communication among students, phenomena which we had been assigning to another part-model. Since "crocks" were also disliked because they gave the student no opportunity to assume medical responsibility, we were able to connect this aspect of the student-patient relationship with still another tentative model of the value system and hierarchical organization of the school, in which medical responsibility plays an important role.

Again, it should be noted that analysis of this kind is carried on in the field as time permits. Since the construction of a model is the analytic operation most closely related to the observer's techniques and interests he usually spends a great deal of time thinking about these problems. But he is usually unable to be as systematic as he would like until he reaches the final stage of analysis.

---

[13] Note again the resemblance to analytic induction.

[14] See Alfred Lindesmith's discussion of this principle in "Comment on W. S. Robinson's 'The Logical Structure of Analytic Induction,'" *American Sociological Review,* 17 (August, 1952), pp. 492–493.

FINAL ANALYSIS AND THE PRESENTATION OF RESULTS

The final systematic analysis, carried on after the field work is completed, consists of rechecking and rebuilding models as carefully and with as many safeguards as the data will allow. For instance, in checking the accuracy of statements about the frequency and distribution of events, the researcher can index and arrange his material so that every item of information is accessible and taken account of in assessing the accuracy of any given conclusion. He can profit from the observation of Lazarsfeld and Barton that the "analysis of 'quasi-statistical data' can probably be made more systematic than it has been in the past, if the logical structure of quantitative research at least is kept in mind to give general warnings and directions to the qualitative observer." [15]

An additional criterion for the assessment of this kind of evidence is the state of the observer's conceptualization of the problem at the time the item of evidence was gathered. The observer may have his problem well worked out and be actively looking for evidence to test an hypothesis, or he may not be as yet aware of the problem. The evidential value of items in his field notes will vary accordingly, the basis of consideration being the likelihood of discovering negative cases of the proposition he eventually uses the material to establish. The best evidence may be that gathered in the most unthinking fashion, when the observer has simply recorded the item although it has no place in the system of concepts and hypotheses he is working with at the time, for there might be less bias produced by the wish to substantiate or repudiate a particular idea. On the other hand, a well-formulated hypothesis makes possible a deliberate search for negative cases, particularly when other knowledge suggests likely areas in which to look for such evidence. This kind of search requires advanced conceptualization of the problem, and evidence gathered in this way might carry greater weight for certain kinds of conclusions. Both procedures are relevant at different stages of the research.

In the post field work stage of analysis, the observer carries on the model building operation more systematically. He considers the character of his conclusions and decides on the kind of evidence that might cause their rejection, deriving further tests by deducing logical consequences and ascertaining whether or not the data support the deductions. He considers reasonable alternative hypotheses and whether or not the evidence refutes them.[16] Finally, he completes the job of establishing interconnections between partial models so as to achieve an overall synthesis incorporating all conclusions.

After completing the analysis, the observer faces the knotty problem of how to present his conclusions and the evidence for them. Readers of qualitative research reports commonly and justifiably complain that they are told little or nothing about the evidence for conclusions or the operations by which the evidence has been assessed. A more adequate presentation of the data, of the research operations, and of the researcher's inferences may help to meet this problem.

But qualitative data and analytic procedures, in contrast to quantitative ones, are difficult to present adequately. Statistical data can be summarized in tables, and descriptive measures of various kinds and the methods by which they are handled can often be accurately reported in the space required to print a formula. This is so in part because the methods have been systematized so that they can be referred to in this shorthand fashion and in part because the data have been collected for a fixed, usually small, number of categories—the presentation of data need be nothing more than a report of the number of cases to be found in each category.

The data of participant observation do not lend themselves to such ready summary.

---

[15] "Some Functions of Qualitative Analysis . . .," *op. cit.*, p. 348.

[16] One method of doing this, particularly adapted to testing discrete hypotheses about change in individuals or small social units (though not in principle limited to this application), is "The Technique of Discerning," described by Mirra Komarovsky in Paul F. Lazarsfeld and Morris Rosenberg, editors, *The Language of Social Research*, Glencoe, Ill.: Free Press, 1955, pp. 449–457. See also the careful discussion of alternative hypotheses and the use of deduced consequences as further proof in Lindesmith, *Opiate Addiction, passim.*

They frequently consist of many different kinds of observations which cannot be simply categorized and counted without losing some of their value as evidence—for, as we have seen, many points need to be taken into account in putting each datum to use. Yet it is clearly out of the question to publish all the evidence. Nor is it any solution, as Kluckhohn has suggested for the similar problem of presenting life history materials,[17] to publish a short version and to make available the entire set of materials on microfilm or in some other inexpensive way; this ignores the problem of how to present *proof.*

In working over the material on the medical school study a possible solution to this problem, with which we are experimenting, is a description of the natural history of our conclusions, presenting the evidence as it came to the attention of the observer during the successive stages of his conceptualization of the problem. The term "natural history" implies not the presentation of every datum, but only the characteristic forms data took at each stage of the research. This involves description of the form that data took and any significant exceptions, taking account of the canons discussed above, in presenting the various statements of findings and the inferences and conclusions drawn from them. In this way, evidence is assessed as the substantive analysis is presented. The reader would be able, if this method were used, to follow the details of the analysis and to see how and on what basis any conclusion was reached. This would give the reader, as do present modes of statistical presentation, opportunity to make his own judgment as to the adequacy of the proof and the degree of confidence to be assigned the conclusion.

CONCLUSION

I have tried to describe the analytic field work characteristic of participant observation, first, in order to bring out the fact that the technique consists of something more than merely immersing oneself in data and "having insights". The discussion may also serve to stimulate those who work with this and similar techniques to attempt greater formalization and systematization of the various operations they use, in order that qualitative research may become more a "scientific" and less an "artistic" kind of endeavor. Finally, I have proposed that new modes of reporting results be introduced, so that the reader is given greater access to the data and procedures on which conclusions are based.

---

[17] Gottschalk, Kluckhohn, and Angell, *op. cit.*, pp. 150–156.

ADVOCATES AND TRUTH-SEEKERS

The following two readings illustrate what we believe is an important difference between law and social science in the task of fact gathering. The McCormick selection sets out a number of clues on how to elicit truthful testimony from witnesses. For example, Professor McCormick advises that free narrative is more conducive to truthful fact finding than is the question and answer technique. Further, he suggests that lawyers ought to steer clear of leading questions, avoid argumentative questions, and refresh past recollections. All of this is important, he says, because it makes the "process of proof look more to the ascertainment of truth and less to the technique of . . . admissibility."

While we believe that both law and social science methodology are designed to unveil the truth, we also think that the relationship between fact finding and the truth is seen quite differently by lawyers and social scientists. Lawyers are advocates by profession. They are employed to establish facts which show a version of the truth most favorable to their client. This is not to say that the legal profession is dedicated to distorting the truth, but rather that it is interested in only a segment of the truth. For example, Mr. X has been given a speeding ticket. His lawyer's task at this point is fairly well established. He must elicit and establish facts which show a segment of the truth which will relieve Mr. X from having to pay the fine. Fact finding, in other words, is almost certain to become a partisan task, and even though Mr. X may have been speeding, the advocate will attempt to throw doubt on the idea. Consequently, he is very likely, as indicated in the reading from *Trials*, to "exercise caution in not asking open-end questions permitting the police witness to volunteer damaging testimony."

The social scientist, on the other hand, is interested in facts in order to display what Bronowski called "the links which give society its coherence." His interest in the truth is to show that there exist relationships between characteristics, which in turn gives him an insight into behavior. Fact gathering, i.e., data collection, then becomes an exercise in attempting to establish what he believes to be the cause and effect pattern. In short, his interest in the truth is non-partisan. He has no client to represent. Facts are indicators which demonstrate whether his hypothesis is true or false.

Excerpts from Chapter 2 (pp. 8-18) of McCormick on Evidence (Hornbook Series) 1954, and are reproduced here with the permission of West Publishing Company, the copyright owner.

Reprinted from 5 American Jurisprudence Trials, Excluding Legally Obtained Evidence, pages 454-458. Published by Bancroft-Whitney Company, San Francisco, California: Lawyers Co-operative Publishing Company, Rochester, New York. Copyright, 1966.

Charles T. McCormick, EVIDENCE
(New York: West Publishing Company, 1954)
pp. 8-18.

The Form of Questions: (a) Questions Calling for a Free Narrative Versus Specific Questions.

The art of direct examination, of telling a composite story from the mouths of your own witnesses, is far more important, though perhaps less difficult, than the area of cross examination. One of the problems of tactics is whether the information which the particular witness will give can best be elicited by a succession of questions about specific facts and happenings or will be brought out more effectively by a general question.

The Form of Questions: (b) Leading Questions.

We have discussed in the preceding section the advantages and disadvantages of the free and unguided narrative of the witness. The danger of the opposite method, the method of drawing out the testimony by specific questions, is the danger that the witness may acquiesce in a false suggestion. The suggestion itself may plant the belief in its truth. The psychologists have verified the convictions of the judges that this danger is greater than one who has had no experience with trials would suppose.

A leading question then is one which suggests to the witness the answer desired by the examiner. One who seeks to prove a set of facts instinctively shapes his questions in this suggestive fashion, and the first task of the beginning advocate is to harness his tongue to the neutral form. Nor is it wholly a matter of form. The question which contains a phrase like "Did he not ?" is obviously and invariably leading, but almost any other type of question may be leading or not, dependent upon the context. It is sometimes supposed that a question which can be answered yes or no is by that fact marked leading, and the beginner finds it helpful to couch his inquiries in the form of a neutral alternative ("State whether or not . . .") to escape the charge of leading, but quite often the former kind of question will not be leading and equally often the latter kind will be. The whole issue is, would an ordinary man get the impression that the questioner desired one answer rather than another? The course of the previous questioning may combine with this question to indicate the desire, but the most important circumstance is the extent of the particularity of the question itself. If the question describes an incident in detail and asks if this happened, the natural inference is that the questioner expects an affirmative answer. Or if one alternative branch of the question is concrete and detailed and the other vague ("Was the sound like the scream of a woman in fear or was it otherwise?") the impression is that the first alternative is suggested. On the other hand, if the question is sufficiently neutral ("At what time did this occur?") or sufficiently balanced ("Was the water hot or cold?") it is not leading. Obviously, a question may suggest a subject or topic, as distinguished from an answer, without offending.

As we have seen, the normal practice is for the careful lawyer to interview in advance all witnesses whom he expects to call for direct examination to prove his own case. This practice is entirely proper, but it does create a probability that the lawyer and the witness will have reached an entente which will make the witness especially susceptible to suggestions from the lawyer as to what the facts were. On the other hand, normally when counsel cross-examines a witness called by the adversary, he has had no opportunity to talk to the witness before, and there is no likelihood that any friendly understanding between them about the facts has been reached. Hence the practice: the judge will ordinarily, if objection is made, forbid leading questions on direct examination; he will ordinarily permit them on cross-examination. But the entire matter of the allowability of leading questions is discretionary, and the judge's action will nowadays not be reviewed unless it is charged that it amounted to, or contributed to, the denial of a fair trial.

The Form of Questions: (c) Misleading and Argumentative Questions.

The examiner may not ask a question which calls for no new fact but merely invokes the witness's assent to the questioner's inferences from or interpretations of the facts proved or assumed. Such a question is subject to objection as "argumentative" but the trial court has a wide range of discretion in enforcing the rule, particularly on cross-examination, the more frequent occasion for such questions. A still more common vice is for the examiner so to couch the question that it assumes as true matters to which the witness has not testified, and which are in dispute between the parties. The danger here is two-fold. If the examiner is putting the question to a friendly witness, the recitation of the assumed fact may suggest the desired answer, and secondly, whether the witness is friendly or hostile, the answer is likely to be misleading. Oftentimes, the question will be so separate from the assumption that if the witness answers the question without mentioning the assumption, you cannot tell whether he ignored the assumption or affirmed it.

Refreshing Recollection

It is abundantly clear from every-day observation that the latent memory of an experience may be revived by an image seen, or a statement read or heard. It is a part of the group of phenomena which the classical psychologists have called the law of association. The recall of any part of a past experience tends to bring with it the other parts that were in the same field of awareness, and a new experience tends to stimulate the recall of other like experiences. The effect of a reminder, encountered in reading a newspaper or in the conversation of a friend, which gives us the sensation of recognizing as familiar some happening which we had forgotten, and prompts our memory to bring back associated experiences, is a familiar process.

As we have seen, the interviewing of witnesses by counsel who will examine them in court is a necessary step in preparing for trial. It is at this stage that the memory of the witness can best be refreshed about the facts of the case, by giving him the opportunity to read his own written statements previously made, or the letters, maps, or other documents in the case. It is only when this review of the data is insufficient to enable the witness to recall the facts while testifying that refreshing his memory on the stand is advisable. If it is matter which a jury would suppose he should remember unaided, the use of a crutch lessons their confidence in the testimony.

Nevertheless, the practice has long been established that in interrogating a witness counsel may hand him a memorandum to inspect for the purpose of "refreshing his recollection," and that when he speaks from a memory thus revived, his testimony is what he says, not the writing. This is the process of refreshing recollection, in the strict and accurate sense. But when this simple but helpful expedient had become established, it was natural for counsel to seek to carry it a step further. If the witness, being shown the writing, states that his memory revived thereby, he may testify, as we have seen, from his refreshed recollection. But it may happen that the witness cannot go so far. He may say that, on looking at the writing, he recognizes it as a memorandum made by him when the facts were fresh in his mind, and therefore, though he has no present memory of the transaction described, he is willing to testify that the facts were as recited in the memorandum.

It is true that any kind of stimulus, "a song, or a face, or a newspaper item," may produce the "flash" of recognition, the feeling that "it all comes back to me now." But the genuineness of the feeling is no guarantee of the correctness of the image recalled. The danger that the mind will "remember" something that never happened is at least as great here as in the case of leading questions. "The problem is complicated by the deceptive certainty of the recognizer. This certainty is a direct function of the similarity of the material. As a result it has an eccentric relation to objective accuracy . . . . It will be objected that, although the foregoing criteria may be interesting as far as recognition is concerned, a present refreshed recollection is based only in part on recognition.

The line between using the writing as an aid to memory and basing one's testimony upon it as a correct record of past memory is sometimes shadowy. Must it be shown that the witness has no present recollection of the matters embodied in the memorandum before he can use it as an aid to memory?" It is usually said that this must appear, but it is believed that this requirement is unsound. The witness may believe that he remembers completely but on looking at the memorandum he would be caused to recall additional facts. As the Chinese proverb has it, "The palest ink is clearer than the best memory." On the other hand, there is here the ever present danger that a suggestive witness may think that he remembers a fact because he reads it. It seems eminently a matter for discretion, rather than rule. Similarly, it would seem that a witness may recognize from present memory the correctness of successive facts set out in a memorandum, but that he may be unable, despite his recognition, to detail those facts from memory without continuing to consult the writing. Accordingly, the statement that a witness once refreshed must speak independently of the writing seems too inflexible, and it is believed that the matter is discretionary and that the trial judge may properly permit the witness to consult the memorandum as he speaks, especially where it is so lengthy and detailed that even a fresh memory would be unable to recite all the items unaided.

"Cross-Examination of Police Officers," Trials, Vol. V, pp. 454-458.

§ 84. — CROSS-EXAMINATION OF POLICE OFFICERS

During the course of the hearing, defense counsel will frequently be confronted with the task of cross-examining the police witnesses. On direct examination, the prosecution has undoubtedly established the elements of probable cause, consent, or some other justification for the search and seizure. Accordingly, defense counsel must attack the credibility of the police witnesses' story on cross-examination if he is to prevent an adverse ruling on the motion to suppress. Having the advantage of listening to defense witnesses' account of the facts in advance of his testimony, the police witness, motivated by the human need for self-justification together with righteous contempt for the decision in Mapp v Ohio,[2] may spin a web of fantasy concerning some justification (usually probable cause or consent) for an otherwise illegal search and seizure.

§ 85. —— AS TO PROBABLE CAUSE; SAMPLE

A sample series of questions follows, showing cross-examination of a police witness as to probable cause.

▣ The reader should assume that the officer has testified to the following facts on direct examination: At about 8 p.m., he and his partner received a radio message in his squad car that there had been an attempted burglary at a particular address a few minutes earlier. He arrived at the address three minutes after hearing the radio message and saw the defendant approximately 50 yards away. As the squad car approached the defendant, the defendant looked at the policemen furtively and started walking rapidly in the opposite direction. The defendant got into a car parked at the curb, started the engine and pulled away, and traveled about half a block down the street where the squad car intercepted him and forced the automobile to the curb. Two television sets were visible in the back seat of the car. The officers checked the serial numbers of the television sets and found that they had been stolen from a warehouse earlier that day. This prosecution arises out of that theft.

Q. Do you know who furnished the information that resulted in the radio message to your squad car?

---

[2] Mapp v Ohio (1961) 367 US 643, 6 L ed 2d 1081, 81 S Ct 1684, 84 ALR2d 933.

A. Yes, it was Mrs. Smith, the next door neighbor to the premises we were told to investigate.

> ■ The source of the original information should be identified to determine whether police informants or undercover agents played a role in the search.

Q. Were you alone?

A. No, I was with my partner, Al Jones. He was the driver.

Q. Were you familiar with the neighborhood?

A. Not really.

Q. It was already dark at this time, was it not?

A. Yes.

Q. Where were the street lights located in relation to the place of the arrest?

A. . . .

Q. Where was the defendant when you first observed him?

A. He was standing on the sidewalk about 50 yards away.

Q. This is a residential area, is it not?

A. Yes.

Q. The defendant then walked away from you?

A. Yes, rapidly.

Q. Had you seen him walking prior to the time that he began to walk away from you?

A. No.

> ■ Defense counsel must exercise caution in not asking open-end questions permitting the police witness to volunteer damaging testimony. For example, after establishing that the police officer never saw the defendant walk prior to the time he walked away, defense counsel might ask, "Then, how do you know he was walking rapidly?" This would be hazardous in that it could elicit the response: "No one walks that fast normally; he was almost running."

Q. You used the term "furtive" to describe the defendant's manner. Could you explain that?

A. Well, . . . suspicious, guilty looking, you know, ...

Q. Exactly what actions of the defendant indicated to you that he looked guilty or suspicious?

A. He opened his mouth as if surprised, and he quickly wheeled about and started to walk away rapidly.

Q. This opening of the mouth and quick turning that you observed were very important to you in determining that you would detain the defendant?

A. Well, in combination with our information already received, they seemed very important.

Q. Since the defendant was dressed in a business suit at 8 p.m. in the evening, without this observation of his furtiveness, you probably would not have stopped his car; is that not correct?

A. Well, his furtive manner was important.

Q. The defendant walked away from you. This also aroused your suspicion, is that right?

A. Yes.

Q. He walked directly toward his car that was parked at the curb?

A. Yes.

> ☐ In the preceding questions, counsel has minimized the significance of the rapid walking and change of direction and general furtive manner of the defendant.

Q. How many other felony arrests did you make that night?

A. This was the only one.

Q. You submitted a report [*or "made a statement," depending on the local police procedure*] in the regular course of police business detailing what you did that night, did you not?

A. Yes.

Q. I request a copy of any statement or report submitted or signed by this witness in connection with the within subject matter.

> ☐ In jurisdictions with statutes similar to the Jencks Act (18 USC §3500) or with similar

procedures for securing witnesses' written statements from the prosecution to test the credibility of testimony, this request will be granted. In other jurisdictions, the motion should be made to establish a record for a Jencks-type test case,[3] but not until the following questions are put.

Q. It is a fact, is it not, that nowhere in any report or statement do you mention the defendant's facial expressions observed when you first noticed him.

A. No, of course not.

Q. Nor did you at any time mention a quick wheeling about?

A. No, I do not think so.

Q. Nor did you ever mention hurried or rapid walking?

A. I do not remember.

Q. You testified earlier that these items were important—in fact, important enough to cause you to make a decision to detain the suspect?

A. Yes, together with the information we had received.

Q. And yet you did not include any of these important factors in any of your written records submitted in connection with this arrest?

A. That is true.

☐ By this mode of questioning, defense counsel has forced the police officer first to identify the elements that he assertedly relied on as indicating probable cause and then to demonstrate the absence of such items in written reports or statements by the officer.

---

[3] Jencks v United States (1957) 353 US 657, 1 L ed 2d 1103, 77 S Ct 1007.

Am Jur:

Discovery of statements of prosecuting and other witnesses. Am Jur 2d, DEPOSITIONS AND DISCOVERY § 317.

ALR Annotations:

Statements and reports of government witnesses producible in federal criminal case under Jencks Act (18 USC § 3500). 5 ALR3d 763.

L ed Annotations:

Validity and construction of Jencks Act (18 USC § 3500) as to defendant's right to production of statements and reports of government witnesses. 5 L ed 2d 1014.

Defendant's burden to show existence of producible document under Jencks Act (18 USC § 3500), giving defendant right to production of statements and reports of government witnesses. 5 L ed 2d 1035.

☐ Further questioning might, for example, elicit testimony as to failure of memory over the months between arrest and trial and testimony placing in question the officer's ability to observe accurately (due to the lighting and the diversion of the officer's attention by the search for the complainant's address in an unfamiliar neighborhood, etc.).

CHAPTER THREE

CLASS NOTES

The following material has been prepared by Professors Audrey and David Franks. We think it is exceptionally clear and straightforward; and it provides the student with some very practical advice on the task of writing questions which all too often possess a deceptive air of simplicity.

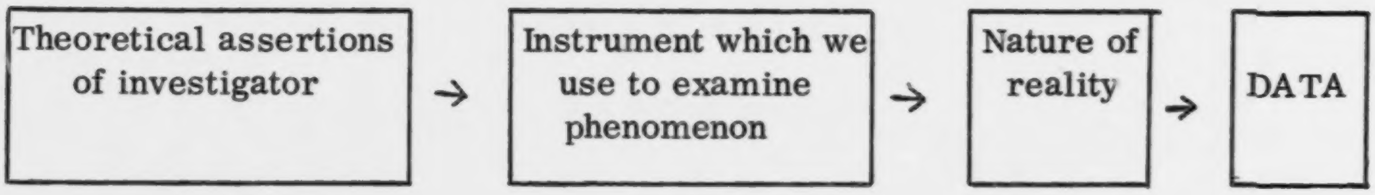
At the end of the notes for this material on the collection of data, we have included four problems. Each involves a decision to be made on the collection of data. Questionnaires, of course, are only one of several possible means, but the issues involved can serve as a starting point for a discussion of how we can move from the theoretical assertions of the investigator to the data of empirical reality.

The Construction of Questionnaires in Sociological Research*

With the exception of the direct observation of social behavior, the questionnaire is probably the social scientist's primary instrument connecting him with the reality he is studying. It is important to see that the questionnaire stands between our theoretical assertions and the reality we are examining. The necessity of seeing these three components of the scientific process as separate things can be underlined by presenting them schematically:

---

*Prepared by Professors Audrey and David Franks.


Theoretical assertions of investigator
Instrument which we use to examine phenomenon
Nature of reality
DATA

When we have these aspects clearly distinguished, we are in a better position to understand how the questionnaire fits into the total research enterprise.

First note that what we refer to as data is not simply the reality we are studying. It is many steps removed from this basic reality and is dependent on how well we have handled the processes designated by the preceding boxes. It is easy to forget this point and accept our data as the unquestioned "foundation of science." The surest way to insure that critical questions will go unasked is to see data as synonomous with reality.

Since data will only be as good as the questionnaire used to collect it, we must be acutely aware of this question: Will our items elicit from the respondent the reality we are interested in, or will they elicit something totally different? If the information we need is collected sloppily, the most advanced techniques of statistics and multivariate analysis are simply wasted motion. Furthermore, mistakes in statistics and analysis can be found and reworked. But once the questionnaire has been administered, mistakes are permanently sealed (and usually hidden) in the remainder of the research process.

The second point that our diagram implies is that good data depends on a questionnaire that has been constructed in the context of carefully worked out theoretical formulations. Science is an enterprise that is built on hypotheses derived from theory. The methodological question we must ask at this stage is does our data contain the information needed to test our hypothesis? Remember we only get what we ask; adequate data means we have asked the right questions. Only well worked out and tightly reasoned theoretical formulations can insure this. It is deceptively easy to overlook the necessity of asking relevant questions, unless one's logical argument and operational definitions are clearly specified before constructing the questionnaire. "Dummy" tables, reflecting exactly how the data will be collected, are an invaluable device for forcing oneself to think through exactly what information must be collected. In sum, a good questionnaire presupposes well-planned theoretical work. Its productivity will only be as good as the theory it reflects.

In analysing what is of critical importance for the construction of questionnaires, there are two basic aspects of the problem which deserve attention. The first has to do with the kind of information being collected. The possible types of information needed can be viewed as forming a continuum, ranging from census type information at one end to items which are conceptually unorthodox at the other extreme. The researcher should have an awareness of approximately where along this line his questions lie, in order to facilitate his general approach to the formulation of items.

Information requests of the first type, those which are census-like in character, present far simpler problems than do their more innovative counterparts. This information is essentially overt rather than hidden and generally conforms fairly closely to common sense usages which are already familiar to the respondents. They are apt to be based on social categories which are widely accepted and easily understood. This kind of data may be used to describe the typical American family (with 2.5 children and 1.2 television sets) or the average income of a variety of occupations in different areas of the country. The questions asked here refer to experiences and units which are well known to the respondent and are the means by which he orders and manipulates his own world.

As a consequence, problems concerning information of this type involve specification of categories known to the respondent beforehand so that he may reply accurately. The questionnaire is well suited to gathering such information and often this may be done both quickly and economically.

On the other hand the information desired may be quite different from the ordinary distinctions and units familiar to the respondent in his everyday life. This kind of information is likely to involve taking a slightly different perspective from that which ordinarily holds. Such items involve points which are subtle, complex, and probably intermingled with other kinds of information in both the mind and the experience of the respondent. Jerome Carlin's study, Lawyers' Ethics, includes many items of this type. "When do you feel most like a lawyer?"

In all probability this is not a distinction typically made, at least in a conscious way, by most of the respondents in this study. Answering such a question requires considerably more effort than replying to a census-like question, both in the process of understanding what is requested and in searching one's memory and organizing an answer. The extra measure of effort involved also suggests that motivation of the respondent is likely to be more of a problem in this type of questioning than in asking simpler, more familiar questions and many more opportunities will arise for misunderstandings. Furthermore, the respondent has little basis for testing the accuracy of his interpretation of the question; because the categories being used are not those with which he has had much experience, he is not able to compare them with his own knowledge and feel sure about his replies. Uncertainty like this is very likely to reduce interest in the whole business and perhaps even engender hostility to it.

These questions are subject to a double handicap--the information requested is likely to be based on distinctions which are unfamiliar to the respondent and which are different from or even in opposition to his daily life, and this kind of questioning and its resultant insecurities for the respondent are likely to result in a lack of interest or even antagonism to the questioning process itself. Both of these difficulties can be modified by a skillful interviewer. Such a person can check for accuracy during the course of the questioning and correct any misinterpretations as they occur (or at least correct gross misinterpretations) while at the same time reassuring the respondent that he is doing well, understanding correctly, etc. This serves to shore up motivation throughout the procedure by virtue of the increased feeling of certainty for the respondent and in the added dimension of genuine social interaction included in an interview instead of a paper questionnaire. The precaution of making sure that serious errors in interpretation of the questions do not go unchecked is a valuable one to the researcher as well.

With these comments in mind, it is not surprising to note that complex information, like that being described, is frequently collected by interview rather than questionnaire. (The examples taken from Carlin's study were done this way.) Although there are separate problems in interviewing which are not a part of questionnaires, the construction and ordering of the questionnaires and items are essentially the same and the differences between the two are not of importance here.

A separate problem arises in the utilization of these more complex and subtle questionings. This type of research and the analysis which inevitably accompanies it does not reproduce the experience of the respondents.

Instead it "distorts" that experience and the resultant reality by virtue of interpreting it and placing it in a framework which is not "natural" to it, by looking at it with new and different eyes. This is the kind of research that makes its subjects cry "Foul!" And from their standpoint, they are correct. Their perspective and insights are not utilized and the new ones produce very different conclusions. This is an intrinsic part of such research questions and cannot be eliminated from any genuine analysis. It is helpful, however, if the researcher handles this in a conscious manner, anticipating some unhappy reactions from his subjects and recognizing that his different persepctive (which is not necessarily the "correct" one) must evoke some such reactions.

Whatever the type of questions to be asked, the researcher is completely dependent on the respondent and this is more true of questionnaires than of any other kind of data collecting. This is very different from data gathered by documents, observations, or even experiment. Self-reporting requires a continuing process of cooperation from the subject and continuous effort on his part to understand what is wanted and to give it. Not even experiments are so directly and unremittingly related to a cooperative mood on the part of the respondent.

Once the researcher consciously recognizes that he is dependent on his subjects, and not vice versa, he is several steps ahead in gaining skill in questionnaire construction. The posing of questions can probably best be seen as one special type of social interaction. It is special because there are a number of aspects which are not shared with social interaction in general but it nevertheless can profitably be seen as falling into this generic category. The major differences between ordinary social interaction and the answering of questionnaires stems from the one-sidedness of the latter.

In questionnaires there are neither mutual goals nor mutually exclusive goals--there are only the researcher's goals. The extent to which the two parties are naive or knowledgeable about the aims and methods always varies, usually to a high degree. If the respondent has the same measure of sophistication about the topic and questions as does the researcher, there almost certainly will not be a successful questioning. When the respondent does not understand the larger issues which the researcher intends to confront but rather deals with the items simply as they are presented to him, he has fulfilled part of the requirements for an ideal respondent. In addition to this, the insertion of personal goals by the respondent into the research situation is likely to have a negative effect.

The ideal respondent is comparatively naive, at least in regard to the research goals and techniques, selfless in terms of his personal goals as opposed to those of the researcher, and continuously cooperative in giving precisely the information desired. Clearly, gems like this are going to be difficult to find; more than likely such paragons would be busy with something else when we did locate them. More reasonably, the researcher must take all possible steps to enable his more ordinary respondent to behave as much like the ideal as possible. No respondent is under any obligation to reply to our questions. Even when there are some coercive measures we can use to bring about compliance in completing a questionnaire, we still cannot force anyone to answer truthfully. Therefore, all measures which tend to smooth social interaction, to eliminate the many potential pitfalls, and to increase or create a feeling of contribution on the part of the subordinate party, will be positive factors in questionnaire construction and questionnaire administration.

Like all social interaction, much of the successful asking of questions is based on skillful role-taking. In turn, accurate role-taking depends in part on one's knowledge of the person involved, and one's ability to make predictions as to behavior from that information. The more accurately we can predict the respondent's range of possible responses to an item and feeling it will arouse in him, the better we are able to choose, phrase, and order questionnaire items which will elicit the desired data. The choice of possible alternative responses to questionnaire items is as critical as the choice of the items themselves. If we do not predict possible responses correctly, if we omit one or more which would be the choice of a sizeable portion of our respondents, we have built a serious error into the

461

data.

A related problem which is also tied to role-taking, is the question of the meaning of the items. It is always a revelation to a novice social researcher (and at times to the more practiced one as well) to discover how many interpretations can be made from seemingly highly structured questions. Research is based on the assumption that the stimuli presented to the subjects will be the same for all of them. Ideally, we can only have one meaning for a question for all respondents. In order to interpret the data, this should also be the same meaning that the item has for the researcher. The most important problem, in a sense, is not "What does the question mean?" but "What will the question mean to the respondent?" An answer to this question really requires an empirical base; a fund of experience with or knowledge about prospective respondents is necessary for accurate role-taking in predicting replies. Where this is not available some form of pre-testing is extremely valuable. It allows culling and sharpening poor items and expanding or revising alternative responses. Much research omits pre-testing entirely or performs it in a perfunctory manner. While it obviously is impossible to estimate what losses in information come about because of such omissions, the often radical changes in questionnaires occasioned by pre-testing suggest that they may be of considerable importance.

Specific Criteria For Questionnaire Construction

Turning now to questionnaires which have been used to obtain information for specific empirical studies, it is necessary to consider some specific recommendations on the construction of items and their arrangement in a questionnaire. Most of the criteria listed below are adapted from Pauline Young's discussion of this topic in Scientific Social Surveys and Research (4th Edition, Englewood Cliffs, N. J.: Prentice-Hall) 1966. It is difficult to specify just what makes a good question and what is a defective one. In reality, it is often impossible to know the difference until the question is put to the intended respondent and unless this is done in pre-testing, it is too late.

Here we shall consider several pieces of research. Each had a different goal and for this reason the kinds of questions they ask will have to be evaluated by different criteria. Recognizing that these rules must be applied with some discretion, let us examine some standards for constructing questionnaires.

Language

1. The vocabulary should be kept as simple as is appropriate for the audience. This is not always as easy as it sounds because it is quite possible to over-simplify, thereby becoming offensive. This is probably less of a danger than too complex wording in most cases, because few people will

indicate their lack of understanding and will not answer the question whether they understand its content or not.

2. The syntax should be clear and straightforward. Complex phrasings and grammatical constructions are very likely to lose respondents in the process.

3. Specialized language should be avoided--except, as indicated earlier, when it is appropriately aimed at specialized groups.

4. Questions should be phrased so that any of the possible alternative answers are equally easy to give. The respondent should not, in other words, be pushed to answer in a particular way and should be free to give an answer that might be considered undesirable.

5. The units or categories about which information is being requested should be clearly defined. This may seem obvious but experience indicates that words and phrases are always far more ambiguous to respondents than they are to the researcher. Where specific information is requested, specific definitions must be given.

6. Specific descriptions are preferable to subjective words--such as "good," "bad," "pretty," etc. Again, where the information desired is to be precise rather than general, the categories available to the respondent should be precise. The question should specify not simply large or small, but how large or small.

Frame of Reference

1. When a respondent must answer a question from a particular point of view or about a different period of time, he must be led to that particular point by the questionnaire rather than being left to find it on his own. If he is not taken there, we do not know whether he is attaching the meaning to what we intend.

2. When questions are asked about highly controversial issues, it is better to break the subject matter into its components than to ask general questions. In this way it is possible to pinpoint the areas of agreement and disagreement and thereby gain much more meaningful information.

3. It is often helpful to get some idea of how much attitudes have crystallized as well as indications of their intensity. The two dimensions do not necessarily coincide and it is very important to know whether strong feelings are a reflection of a general perspective which has been held for years or the result of some immediate precipitating situation.

4. The information requested should be within the range of what is available to the respondent; don't ask for information he cannot be expected to have. Sometimes this may be rather difficult to catch. How consistently, for example, do college students know their fathers' incomes? When exact information is not necessary for the researcher's purposes, it is helpful to indicate to the respondent that only an estimate is needed.

5. The level of the questions should be regulated by the anticipated comprehension level of the respondents. This should always be considered very carefully in constructing the final forms of all items.

Arrangement of Questions

1. The first questions should either be quite easy to answer or else designed to arouse interest. Obviously, the important thing is to get the respondent involved in answering the items as quickly as possible. Avoid controversial questions and items which are difficult to answer or which ask for information which may not be readily available to the respondent. These should come only after the respondent is well into replying to the questionnaire.

2. Caution should be used to avoid having earlier questions affect answers to later items. It is generally better to ask questions about the respondent's current situation before asking about past or future. Other time periods may blur or distort the present. Where the researcher is not certain about the effects of the sequence, it is wise to pre-test several forms of the questionnaire.

3. Questions pertaining to a particular period of time should be grouped together to facilitate answering. Do not require a respondent to shift mentally from one time period to another and back again.

4. Subject matters should also be grouped together so that as few jumps in focus as possible are necessary. Where one must choose between grouping time periods together or grouping subject matters together, it is usually best to keep subjects together.

General

1. Probably the most important issue to raise about each item on any questionnaire is: IS THIS QUESTION NECESSARY? The novice researcher particularly, tends to be over-enthusiastic about data collecting and tends to ask too many questions. Too many questions mean unanswered questions. On the other hand, once the items have been chosen and the questionnaires distributed, no further information will be gathered so it is of critical importance to consider exactly what information will be needed and to be certain it is being collected. A vast number of questions, however, does not in itself guarantee that the desired information will be available; that requires careful thought and planning.

2. Would several questions on a particular subject be more effective than a single one? In dealing with complex or controversial topics, the researcher should guard against collapsing too much information into one item.

3. Should some questions be phrased in more concrete terms, more directly relevant to the respondent's experience? Abstract questions are more difficult to answer and less accurate.

4. Do any questions contain unstated assumptions which will affect the way the question is answered? If so, these should be made as explicit as possible. Otherwise the question cannot be the same for all respondents but rather will be different depending on their knowledge of these assumptions.

5. Can any of the wording of any items be objectionable? Obviously this is difficult or impossible to assess with certainty, but it must be carefully considered. A respondent who is offended by an item is partially or totally lost to the research.

Examples of Questionnaire Problems

Listing of criteria, such as was done above, is interesting but not really very relevant until it can be applied against some particular items. The most meaningful application is to questionnaires actually utilized in social research, rather than hypothetical cases. Items from three separate questionnaires will be examined in regard to the questionnaire criteria which have been delineated. Each questionnaire was used in a very different way from the others as part of separate pieces of research in the sociology of law. The differing intents of the research projects and the different audiences being addressed will be evident in the kinds and forms of questions asked.

The shortest and simplest questionnaire of this group was used to collect information from all members of the Colorado Bar Association concerning the economics of the legal profession. The questionnaire was mailed and drew a 46% return. Because of the kind of respondents and the fact that a great deal was known about them before these data were collected, a large number of assumptions could be made about them and their background. Nevertheless, definitions of five critical categories are given before any questionnaire items begin; e.g., "(a) A 'sole practitioner' is one who practices law with no 'partners', although he may employ one or more 'associates.'" Thus, assumptions about a prior definitional consensus are omitted almost entirely and terms which are of prime importance are clearly specified. (See #5 in Language and #4 in General above.)

The first four questions are quite simple and easily answered by the respondent; they fulfill very well the first rule in the Arrangement section. The respondent almost certainly will have no difficulty completing these items and by then will be well into the questionnaire.

1. In what judicial district do you maintain your principal office?
2. What is the population of the community in which you maintain your principal office? (followed by seven population categories as choices)
3. How long ago were you first admitted to the bar of any state? Answer to the nearest full year. (followed by eight categories of years - under 1 year, 1 to 5 years, etc.)

4. How long ago were you first admitted to the Bar of Colorado? (same categories)

In addition to being simple and straightforward, these questions also set the stage for the next group of items, which focuses on more specific economic aspects of the respondent's law practice. By the time the first four have been answered, the respondent is oriented to his present situation and its history. He has been led through some necessary steps to the frame of reference from which he can answer the next question. (See #1 in Frames of Reference.)

Two items are particulary good examples of careful phrasing on issues which might be delicate or controversial. Considerable effort has been made to avoid any potentially offensive wording.

8. This economic survey can provide much more information if it is possible to analyse the lawyers' incomes in light of the particular community in which they work. We wish to stress again that these answers will be kept completely confidential and the anonymity of the respondents will be protected. If you do not feel that it is an intrusion, will you please fill in the name of the community in which you practice on the space provided.

9. To the best of your recollection, where did you stand in your law school graduation class?
    1. Top 25%
    2. Second 25%
    3. Third 25%
    4. Bottom 25%
    5. Did not graduate from law school
    6. Don't remember

Each of these questions has been phrased on the basis of extensive role-taking, imagining the potential reactions to the presentation. The respondent is assured, in each case, that the question is not intended to embarrass him or put him on the spot. The words suggest clearly that it would be possible *not* to cooperate for very justifiable reasons. This acceptance of a negative reaction to the researcher's desires blunts the intensity of that reaction and makes it easier for the respondent to cooperate.

Controversial issues are better handled by being broken down into their components, rather than being dealt with globally (See #2 in *Frame of Reference*). This questionnaire includes a good example of how this can be done.

12. How do you feel about a program extending legal services to the poor:
    a. Which uses federal funds?
        1. Very favorably inclined
        2. Somewhat favorably inclined
        3. Somewhat unfavorably inclined
        4. Very unfavorably inclined

    b. Which uses municipal, county, or state funds?
        1. Very favorably inclined
        2. Somewhat favorably inclined
        3. Somewhat unfavorably inclined
        4. Very unfavorably inclined

    c. Which uses charitable funds such as those of United Ways, Community Chest, or similar organizations?
        1. Very favorably inclined
        2. Somewhat favorably inclined
        3. Somewhat unfavorably inclined
        4. Very unfavorably inclined

    d. Which uses the services of unpaid, volunteer lawyers?
        1. Very favorably inclined
        2. Somewhat favorably inclined
        3. Somewhat unfavorably inclined
        4. Very unfavorably inclined

Imagine what meager information would be the result had the question included only the general issue of extending legal services to the poor. By specifying components of the issue and asking about each of them in turn, we gain great precision as well as richer detail.

The second piece of research and data collecting to be examined here also concerns lawyers and their practice but in a very different way. This study was reported as Lawyers' Ethics, by Jerome Carlin (New York: Russell Sage Foundation) 1966. The data for this study was gathered by interview rather than questionnaires; that is, the subjects were asked questions by an interviewer, in direct interaction, rather than filling out a printed form to give their responses. Since the interview schedule was a highly structured one in which the interviewer asked questions which were written out in detail and then checked off the appropriate answers, it is quite comparable to questionnaires. Apparently the reason for using interviews rather than questionnaires in this study lay in the extensive amount of detailed data which was needed and the complexity of modifying the questions according to the work situation of the respondent (different questions were used for lawyers practicing alone, as associates, as employees, as partners and with other law firms in a suite, as partners but without other law firms in the suite, etc.). Furthermore, personal contact facilitates the eliciting of such large amounts of data. As was indicated earlier, when a great deal of time and effort are required of a subject, an interview often is necessary because the actual interaction gives satisfactions and rewards to the respondent which are essential for his continued participation. Lawyers' Ethics is a clear example of this. Because the questions are perfectly appropriate to a questionnaire, they will be helpful examples of problems in questionnaire construction.

The interview begins with two very simple questions which direct attention to the respondent's professional career history:

1. In what year were you first admitted to practice? __________(year)
2. Did you clerk in a law office either before or during law school?

Yes _____
No _____

The complexity of his schedule lies in its making many distinctions between different categories of law practice, and making these distinctions in relation to many other areas. This requires that the units being enumerated are consistently clearly defined and specified. Here is a particularly good example of this kind of question.

3. What was your first job after law school? (Hand respondent Card 1.) (Card contains list of categories below.)

   a. How long were you there? (Indicate duration in appropriate box in chart below.)

   b. What was your next position (Etc. up to and including present position)?

Employee (or associate) of:	1st	2nd	3rd	4th	5th	6th
a. Individual lawyer						
b. Firm with fewer than 5 lawyers						
c. Firm with 5-14 lawyers						
d. Firm with 15 or more lawyers						
Partner in firm with:						
a. Fewer than 5 lawyers						
b. 5-14 lawyers						
c. 15 or more lawyers						

On own:
a. Employing 1 or more lawyers
b. Without lawyer employees

Corporate or union legal department (name)
Government legal department
a. Local (name)
b. State (name)
c. Federal (name)
Nonlawyer (specify)

From the information collected in this item, career development of the respondents can be specified and different kinds of development can be compared in terms of other variables. If the information on the size of law firms had been omitted, no analysis could ever be made from this data concerning effects of size of firm formerly affiliated with on present ethics. Since size turned out to be a critical variable in the overall analysis here, this would have been a serious omission.

Three further examples of questions in which careful specification is made are:

49. Let's consider your individual clients, including those individuals who are principal owners of closely held corporations. Within the past year, what proportion of all the individual clients you have done work for were people for whom you had previously done work, as opposed to people coming to you for the first time?

Previously worked for ________%
First time ________%
(100%)

118. What college(s) did you attend? ____________________
____

If any:
a. How many years did you attend college? ________
b. Did you receive a degree?
Yes ______
No ______

If yes, and attended more than one college:
From which did you receive a degree? ____________________

106. How often, during the past few years, has a question of professional ethics come to your attention under each of the following circumstances? (Card 31)

	Never	Less than Once a Year	Once or Twice a Year	Several Times a Year	One or More Times a Month
a. While reading a legal journal or periodical					
b. During informal discussions with other attorneys					
c. In the course of handling a case or matter					
d. Other (specify)					

Note how carefully, in each of these cases, the exact information desired is specified. This allows the respondent to answer in a way that is likely to satisfy him, since he feels certain that he is answering correctly and not misinterpreting the question, and also gives the necessary data to the researcher.

Leading questions should be avoided if the data are to be accurate. If each alternative is not equally easy to answer our data can not be correct. This is of particular importance when dealing with topics that are controversial or which the respondent may feel reflect on himself, his status, his sense of personal worth. The following question *could* be interpreted as questioning one's personal competence, but has been phrased in such a way that this potential is minimized.

Pll. In important matters that you handle, do you generally take full responsibility for setting the fee, or do you review it with your partner?

On Own ________
Review ________

Another important aspect of the use of language in item construction concerns the use of specific description rather than more general, subjective words. Question 43 demonstrates one way in which this approach can be implemented.

43. What proportions of your business clients fall into the various categories under each of the following headings (Card 10):

Legal Form of Enterprise	
Individual proprietorship	_____%
Partnership	_____%
Closely held corporation fewer than 10 shareholders	_____%
Public corporation	_____%
	(100%)

Net Worth of Enterprise	
Under $20,000	_____%
$20,000 to $50,000	_____%
$50,000 to $100,000	_____%
$100,000 to $500,000	_____%
$500,000 and over	_____%
	(100%)

a. What proportion of your business clients have you represented more or less continuously for a year or more? __________%

b. What proportion have you represented more or less continuously for over five years? __________%

For each of these inquiries, percentages are likely to lead to more useful data than would subjective descriptions of "most," "few," "many." This framework encourages the respondent to think about the topic as a whole and to relate the parts to that whole rather than reacting to each segment separately. The categories are very clearly defined and probably are fairly easy for the respondent to react to.

Frequently it is difficult to specify precisely what referent is meant in an item, especially when distinctions have been made along several dimensions. This can lead to awkwardly phrased and constructed questions. One example is:

> P7. Over the past 12 months, roughly how many hours a week on the average have you spent in law practice on matters not connected with the firm, excluding unpaid charitable service?
>
> __________ per week

This question could be better worded by breaking the ideas down into two sentences, the first describing a portion of the distinctions to be made (over past 12 months average number of hours per week, not connected with the firm, unpaid charitable service). The second would complete the definition of precisely what is being discussed and asking for the information. Where a very specific piece of information is being requested it is essential that the respondent understand exactly which of several possible categories is actually under discussion, it is often preferable to structure the question into two or even three sentences, rather than trying to get so much information into one.

This problem of getting the respondent to look at just what the researcher wants him to look at lies in the realm of frame of reference. The first criterion listed earlier in that area concerned leading the respondent to that point, rather than simply assuming he can get there. Efforts along these lines may be included in the item itself, as in the question immediately preceding this one, or may be presented as introductory material before the questions are asked. An example of the latter is:

> EMPLOYEES: I want to ask you some questions about your participation in these activities with your employer, and fellow employees if you have any. Please note that in the following questions the phrase "another lawyer in your firm" refers to your employer, and fellow employees if any.

In this presentation the respondent is clearly directed to the material desired and the proper frame of reference is established for him.

Questions which deal with controversial or sensitive issues should be separated into their components as much as possible. Even in otherwise good questionnaire construction, some lapses in this can slip by. The following question probably asks for information in too global a way to be really useful.

27. Have you recently considered leaving this suite? Yes _____
No _____

If yes, what have you considered doing?

Compare that phrasing with the precision of response possible in the following item.

77. a. How would you rank the following occupational groups with respect to how important a contribution they make to society? (Card 14. 1 = Most Important, etc.)

b. How does the public rank these occupations in terms of prestige? (1 = Most Prestige, etc.)

Contribution	Rank (1-5)	Prestige	Rank (1-5)
Teachers		Teachers	
Engineers		Engineers	
Lawyers		Lawyers	
Businessmen		Businessmen	
Doctors		Doctors	

Each of the items raising questions concerning professional ethics also examines several components of these complex situations. Here, of course, the data is very critical to the basic analysis of the study.

96. Lawyer A is negotiating a personal injury claim with an insurance company adjuster. The adjuster, with whom A has had previous dealings, is under pressure from his supervisors to hold down the amount of recovery of this particular claim and indicates this fact to A, saying that if A will cooperate this time he (the adjuster) will be able to take care of him next time. The offer is in striking distance of a fair amount. (Card 22)

a. How often in the past five years has a situation like this come up in your practice? Never ______ Sometimes ______ Often _____

b. If ever:
What have you done? ____
If never:
If it came up, what would you do? ____________________________

c. Suppose A decides to go along with the adjuster, would you
Approve ________
Disapprove ________
Neither Approve Nor Disapprove ________

The last two items yield data that can be handled in a variety of ways rather than being restricted to one, and allow comparisons of various sorts to be made. In the ethics questions, the separation of behavior and approval of that behavior into two items (rather than assuming that behavior reported would also be approved by the respondent) allowed the author to make some highly perceptive analyses of his data.

An assumption which is sometimes made in research is that all attitudes are equally crystallized. That is, all beliefs and evaluations are held with approximately equivalent intensity and commitment to them. The differences may, in fact, be more important than differences in content. Here is an item which attempts to get at this point.

29. a. Which of the following characteristics would you most want to know about another lawyer before deciding to share office space with him? Which would you want to know about next? Which next? (Card 5: Rank only 1, 2, 3.) Which would you not be concerned about at all? (Check)

	Office-Sharer	Partner
Type of practice		
Personality		
Loyalty to Clients		
Competence		
Fairness in dealing with colleagues		
Honesty in dealing with officials		
Business-getting ability		

b. Which of these characteristics would you most want to know about another lawyer before going into partnership with him? Which would you want to know about next? Which next? (Rank only 1, 2, 3.) Which would you not be concerned about at all? (Check above.)

Presumably, values which apply to partners and not to office-sharers are more highly crystallized than those which apply to neither, but not as crystallized as those which apply to both possibilities. In this case, also, the data elicited by such a question is much more flexible and amenable to more kinds of data analysis than would be the case in a narrower and more simplistic kind of question.

To say that questions should only be asked about information which lies within the scope of the respondent sounds very obvious. In looking at questionnaire items, however, it is evident that criteria for this point are easy to forget. This schedule, for example, contains three items which are of questionable appropriateness in this regard.

50. Of all the individual clients you have done work for in the past year, roughly what proportion are (Card II):
People earning under $5,000 a year _____%
People earning from $5,000 to $10,000 a year _____%
People earning from $10,000 to $20,000 a year _____%
People earning over $20,000 a year _____%
(100%)

53. What proportion of your individual clients are:
Jewish ___________%
Catholic ___________%
Protestant _________%
(100%)

121. What was your father's approximate income at that time (when respondent entered law school) or before, if deceased then? (Card 32)

Under $1,000
$1,000 to $2,499
$2,500 to $4,999
$5,000 to $6,999
$7,000 to $8,999
$9,000 to $11,999
$12,000 to $14,999
$15,000 to $19,999
$20,000 to $29,999
$30,000 and over

Certainly there are likely to be many respondents who would be unable to answer such questions with any appreciable accuracy. Such items not only are wasteful of the respondent's time and the researcher's time, they also lead to poor and/or incorrect analyses and to frustration and dissatisfaction on the part of the respondent. More sensitive efforts at role-taking and careful pre-testing should eliminate the bulk of such items.

The arrangement of specific items into the actual questionnaire is influenced by a number of factors, some of which have already been indicated (the initial items should be simple and easily answered). The grouping of questions about particular subjects and/or time periods is fairly clear-cut and not difficult to implement. Avoiding contamination of response to one question because of prior reaction to another is, however, a much more subtle problem. It is probably well to have several observers examine the questionnaire from this standpoint.

Carlin's interview schedule contains one especially good example of careful placing of items. Following the 13 situations describing ethical conflicts, there are two items concerning knowledge about the official canons of ethics which apply to the situation. (The two issues mentioned were involved in the preceding ethical items.)

104. a. Do you happen to know what the canons or opinions dealing with the canons have to say about accepting commissions on title insurance policies? Yes _____
No _____

If yes:
What is the official position? ____________________

b. Do you happen to know what the canons or opinions dealing with the canons have to say about representing a party in a controversy with a previous client? Yes _____
No _____

If yes:
What is the official position? ____________________

A question such as this must come after those asking "What did you do?", if we are to get any accurate information. Alerting the respondent to the officially correct answer and reminding him of it can only put additional pressure on him to give this answer and no other. Asking such a question after he is asked what he did, in fact, may make him uncomfortable but it reduces the risk that he will tell us what he thinks he should have done rather than what he did do. Sensitivity to such problems is highly advantageous at all stages of questionnaire construction.

The third research questionnaire to consider in relation to general criteria of question construction is a part of a much larger study by Isidor Chein, Donald L. Gerard, Robert S. Lee, and Eva Rosenfeld, reported as *The Road to H*, (New York: Basic Books) 1964. The use which was made of the questionnaire in this study was markedly different from the questionnaires discussed above, and from most instruments utilized by sociologists. Some items are directed toward the overt information which is being requested, but other items are intended and used in a different way. Rather than asking a particular question in order to collect specific information about the respondent, these items present a series of stimuli to which the respondent is asked to react, and his reactions, along with those of many other subjects, are examined for certain patterns of response. The researcher here is not so much concerned with accumulating specific content about his respondents as with discovering how certain *categories* of persons respond to certain *categories* of stimuli. From these responses, he may (as did Chein, *et al.*) derive information indirectly concerning the specific content of the respondents' values, beliefs, or experiences. This indirect route is used when direct questioning on the subject would be difficult or impossible.

Portions of the questionnaire, of course, are items concerning basic information about the respondent, necessary to categorize him along important dimensions. Initially he is asked for birth date, sex, school, and class in school, but not his name. As the researchers (and one of the respondents) realized, the requested information would be sufficient to identify the respondents so as to put two different questionnaires together to get information regarding ethnicity from teachers. It was believed by the researchers that questions on ethnic background were too sensitive to ask directly and this method was devised to get the data deemed necessary. One respondent (all were eighth grade students in the New York City neighborhoods under study) realized that he could be recognized by the questionnaire information, and tore up his paper before it could be collected. This points up a frequent hazard of such research; direct questions may offend but indirect items may offend more. In this case, most of the desired information was collected. Some questions can be raised about both the ethics of such techniques and their long range utility. If sizeable portions of the population become antagonized by certain specific practices of social research, the techniques may well be questionable even on purely pragmatic grounds.

The arrangement of the attitudinal items might well have been improved. The first two items are:

1. I am a very lucky person.
   ____ I agree
   ____ I do not agree

2. I often think that parents don't want their kids to have any fun.
   ____ I agree
   ____ I disagree

Each of these items may well be disturbing to some respondents and a negative reaction to them might evoke considerable caution in responses to the remainder of the questions. Several other items in the same section seem better suited to begin the questionnaire.

3. You should never be loud around the house.
_____ I agree
_____ I disagree

26. Even if you can't stand some people, you should still be nice to them.
_____ I agree
_____ I do not agree

Items such as the latter two would be very unlikely to tap into deep-rooted fears, excessively poor parent-child relationships, etc.

Since this questionnaire was designed for eighth grade boys who live in areas of moderately high to very high delinquency, the exceptions as to their academic level were not high. Considerable effort was exerted to insure that the content of the items would not exceed the comprehension level of the respondents. Some indications of this effort are clear in items such as the following:

9. Who have fewer brains?
_____ fellows who use heroin
_____ fellows who do not use heroin
_____ both about the same

10. Who are more fun to be with?
_____ fellows who use heroin
_____ fellows who do not use heroin
_____ both about the same

Simplicity of vocabulary and syntax are demonstrated in the following:

4. Just taking a little heroin once in a while never really hurt anybody.
_____ I agree
_____ I do not agree

7. A person should never take heroin no matter what.
_____ I agree
_____ I do not agree

A number of items in this questionnaire do not specify clearly the subject to which they refer. Not surprisingly, these create difficulties in analysing the data; the meaning of different responses is not clear.

47. Are you the youngest person who lives in your apartment?
_____ Yes
_____ No

48. Are you the only person under eighteen who lives in your apartment?
_____ Yes
_____ No

Either of these might be interpreted as referring to the particular family's quarters or to the apartment building in which the respondent resides. This uncertainty is also evident in the data analysis. Another example is:

50. Did you ever have a chance to use heroin?
_____ Yes
_____ No

In this question, the researchers became aware that a "chance" to use heroin was not actually very specific; it might refer to an open invitation with the drug immediately available in the situation, or it might refer to a discussion about future possibilities. It is not possible to designate how similar the experiences of those who reply "yes" actually are. These three items do not specify the unit of enumeration clearly enough to be usable.

One item might be seen as a leading question, in that it encourages one type of reporting rather than another. It seems to be designed somewhat similarly to the way the original Kinsey studies were; the assumption of some deviance is made and the establishment of conformity is the burden of the respondent. Hence;

49. About how many people do you know who use heroin?
Check one of these.
A. One or two people _____ A
B. Three, four, or five people _____ B
C. Six or more people _____ C
D. I don't know anybody who uses heroin _____ D

The way in which this question is phrased makes it much easier to report knowing someone who uses heroin than reporting no one. It also suggests that knowing a heroin user is more "normal" than not knowing such a person. All of this obviously exerts certain pressures on the respondent to report some familiarity with heroin usage.

One series of items in this questionnaire is designated to get at how much information the respondent has about narcotics. Thus for many respondents these items would be outside of their normal sphere, and in this case this is the point of the questions.

19. Most steady users begin using heroin before they are thirteen years old.

20. It is legal to buy heroin from a drugstore, but a person has to have a doctor's prescription.

30. No city hospital will take in a drug-user for treatment.
(Each question can be answered _____ true
_____ false
_____ I don't know.)

This type of question must be very carefully presented, because attempting to answer questions about which one has insufficient information is extremely frustrating. When the intent of certain items is to ascertain what the level of knowledge is about some topic, considerable care must be exercised to insure some satisfactions for those whose knowledge is minimal. If this is ignored, there will probably be an unusually large number of incomplete questionnaires.

The three questionnaires examined here differed notably both in what they were designed to do and in how well they did it. The first, the economic survey, is the simplest and probably the most effective in attaining its goal. Essentially this is simply a census and it is not difficult to find items which would be appropriate to the task. The test of adolescent attitudes, on the other hand, wanders far from a simple census, and probably encounters some unnecessary difficulties as a result of its content. This may be in part a result of the small role the questionnaire played in the overall research. The data collected by this means was not a critical portion of the larger study. Because of this, it may have received less attention than would have been appropriate for good questionnaire construction. The interview schedule on lawyers' ethics was not only the longest but certainly the most thoughful of the group. It seems reasonable to conclude that this is a function of its being the sole source of data for a large-scale study. In this case, the obvious effort invested in a project contributed very important results.

These three questionnaires seem to make a good case for the general approach to the problem presented initially: good questionnaire construction can only emerge as a function of clear and adequate theory and conceptualization, knowledgeable role-taking with the intended responderts, and a sharp notion of exactly what one intends to do, both generally and in each specific item. When these are given adequate thought, a good and informative questionnaire is likely to be the result.

SOME ILLUSTRATIVE EXERCISES

1. A survey project in a large metropolitan area has reported the following facts:

   A. Approximately 90% of the 500 persons in the sample (white, Anglo-Saxon, male) chose response (1) or (2) to the following item:

   As far as Negroes are concerned:

   (1) I like them

   (2) I like them somewhat

   (3) I dislike them somewhat

   (4) I dislike them

B. Approximately 70% of the persons in the sample reported that "They would be very upset if someone in their family married a Negro." How would you interpret these findings? How would you test your intepretation? Specify the kind of data you would need and how you would collect it.

2. Two social-psychologists, who conducted psychological testing in a prison, report that "the men who are regarded as leaders by the inmates are almost always very hostile to the custodial and administrative staff, aggressive, often violently anti-social, and thoroughly enculturated into a criminal way of life." They also report that it is extremely difficult to establish a relationship of trust with more than a handful of inmates and even this takes a very long time.

   How would you test the assertion of the two social-psychologists?

3. In the development of a federal program for aid to urban areas, it has been argued that small business is a major factor contributing to the civic welfare of a community. There is no doubt that some funds will be provided for aid to small business in the program. The level of aid in this program, however, will depend in part on just how important small business can be in influencing the civic welfare.

   How would you measure the influence of small business?

4. A study has been made of the membership of the American Psychological Association with regard to the ethical issues involved in psychological research. Of the questionnaire mailed to the entire membership, approximately 20% were returned. The results indicated that the greater the age of the respondent, the greater the number of psychological safeguards that were thought to be desirable for those who were the subjects in research projects.

   How would you interpret this finding? How would you test your interpretation?